OM of: US-09-664-641-10 to: SPTREMBL_19:*   out_format : pfs

Date: Jun 4, 2002 5:49 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-O=/cgn2_1/USPTO_spool/US0966464l/runat_04062002_110933_9250/app_query.fasta_1.3697
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641_@CGN1_1_312 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_rodent:Q9Z0W6	+	4059.50	5875.12	0.0	1056	Q9z0w6 mus musculus (mouse) .
sp_human:O15404	+	3841.00	5551.10	0.0	744	O15404 homo sapiens (human). ca
sp_vertebrate:O90W3	+	3425.50	4951.40	1.5e-268	1256	O90wj3 xenopus laevis (african
sp_human:Q96HP2	+	2039.00	2947.67	1.9e-156	391	Q96hp2 homo sapiens (human). un
sp_invertebrate:Q9VTB6	+	852.00	1205.56	4.6e-60	1798	Q9vub6 drosophila melanogast
sp_invertebrate:Q94046	+	524.00	734.07	1.4e-33	1074	Q94046 caenorhabditis elegan
sp_invertebrate:Q9O370	+	524.00	734.05	1.4e-33	1076	Q9u370 caenorhabditis elegan
sp_human:Q14676	+	342.50	463.12	8.9e-19	2089	Q14676 homo sapiens (human). k
sp_human:Q96QC2	+	340.50	460.21	1.3e-18	2090	Q96qc2 homo sapiens (human). k
sp_plant:Q9LVF7	+	316.50	432.68	8.9e-17	1041	Q9lvf7 arabidopsis thaliana (m
sp_rodent:Q91VX3	+	283.50	382.38	4.5e-14	1296	Q91vx3 mus musculus (mouse). s
sp_plant:Q9Z8R18	+	265.50	361.78	1.1e-12	765	Q9zr18 arabidopsis thaliana (m
sp_human:Q9UEB9	+	263.50	352.22	2.0e-12	1435	Q9ueb9 homo sapiens (human). k
sp_human:Q92547	+	263.50	351.40	2.0e-12	1550	Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5	+	229.00	307.74	1.0e-09	838	Q9p5j5 neurospora crassa. cons
sp_invertebrate:Q9VFR7	+	225.50	303.46	1.9e-09	776	Q9vfr7 drosophila melanogaster
sp_invertebrate:Q24163	+	220.50	291.75	5.5e-09	1180	Q24163 drosophila melanogaste
sp_invertebrate:Q9YED3	+	220.50	291.71	5.6e-09	1186	Q9ved3 drosophila melanogast
sp_invertebrate:O44011	+	219.50	288.08	7.2e-09	1457	O44011 dictyostelium discoid
sp_invertebrate:Q9N4E8	+	216.00	293.94	9.6e-09	517	Q9w4e8 drosophila melanogast
sp_plant:Q93VQ3	+	215.50	283.70	1.4e-08	1272	Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853	+	214.00	293.23	1.3e-08	420	O76853 dictyostelium discoide
sp_fungi:P79025	+	214.00	291.75	1.4e-08	483	P79025 emericella nidulans (as
sp_virus:Q98148	+	214.00	282.47	1.9e-08	1162	Q98148 kaposi's sarcoma-associ
sp_invertebrate:Q9VP17	+	213.50	274.60	2.6e-08	2285	Q9vp17 drosophila melanogaster
sp_vertebrate:Q9OYI2	+	212.50	275.54	2.9e-08	1822	Q9oyi2 fugu rubripes (japanese
sp_invertebrate:Q9Y0C9	+	212.00	283.02	2.4e-08	838	Q9y0c9 dictyostelium discoide
sp_invertebrate:O18346	+	210.50	278.87	3.4e-08	1010	O18346 drosophila hydei (fru
sp_virus:Q9QR71	+	210.00	276.96	3.9e-08	1129	Q9qr71 kaposi's sarcoma-associ
sp_invertebrate:Q9VMS2	+	208.50	266.64	6.7e-08	2439	Q9vms2 drosophila melanogast
sp_invertebrate:Q95ZG5	+	207.00	262.29	9.6e-08	2994	Q95zg5 dictyostelium discoid
sp_invertebrate:Q9VC5	+	206.50	273.98	6.9e-08	925	Q9vc5 drosophila melanogaste
sp_invertebrate:Q95YM8	+	206.50	268.20	8.4e-08	1598	Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641	+	205.50	288.88	4.8e-08	197	Q17641 caenorhabditis elegan
sp_invertebrate:Q9VN59	+	205.00	271.70	9.2e-08	934	Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485	+	204.50	274.64	8.9e-08	660	Q94485 dictyostelium discoide
sp_invertebrate:Q9NBT2	-	203.00	255.01	2.1e-07	3439	Q9n8t2 trypanosoma brucei. p
sp_human:O15410	+	202.50	271.86	1.3e-07	652	O15410 homo sapiens (human). ca
sp_human:O75557	+	202.50	259.89	1.9e-07	2023	O75557 homo sapiens (human).
sp_human:Q9UND7	+	202.50	259.89	1.9e-07	2023	Q9und7 homo sapiens (human). c

sp_virus:Q9DUN0	+	201.00	265.41	2.0e-07	976	Q9dun0 kaposi's sarcoma-asso
sp_rodent:O64507	+	200.00	279.24	1.4e-07	230	O64507 mus musculus (mouse).
sp_human:Q9BT44	+	200.00	268.27	2.0e-07	649	Q9bt44 homo sapiens (human).
sp_virus:O40947	+	200.00	262.80	2.5e-07	1089	O40947 kaposi's sarcoma-ass
sp_invertebrate:Q9U9S7	+	200.00	255.74	3.1e-07	2123	Q9u9s7 dictyostelium disc

seq_name: sp_rodent:Q9Z0W6

seq_documentation_block:
ID   Q9Z0W6   PRELIMINARY;   PRT;   1056 AA.
AC   Q9Z0W6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PAX TRANSCRIPTION ACTIVATION DOMAIN INTERACTING PROTEIN PTIP.
GN   PAXIP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB;
RX   MEDLINE=20368635; PubMed=10908331;
RA   Lechner M.S., Levitan I., Dressler G.R.;
RT   "PTIP, a novel BRCT domain-containing protein interacts with Pax2 and
RT   is associated with active chromatin.";
RL   Nucleic Acids Res. 28:2741-2751(2000).
DR   EMBL; AF104261; AAD17923.1; -.
DR   MGD; MGI:1890430; Paxip1.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 6.
DR   SMART; SM00292; BRCT; 5.
DR   PROSITE; PS50172; BRCT; 5.
SQ   SEQUENCE   1056 AA; 119331 MW; 99CAF4EF4EE53ED6 CRC64;

alignment_scores:
          Quality: 4059.50   Length: 1142
            Ratio: 4.572   Gaps: 11
Percent Similarity: 77.758   Percent Identity: 70.928

alignment_block:
US-09-664-641-10 x Q9Z0W6

Align seg 1/1 to: Q9Z0W6 from: 1 to: 1056

        33 CCGGAATTCCCGGTCGACGATTCGTCGAGTGCAAGACGGGCGGGCTGC 82
            |||||| ||| ::: ||||||:::::||:::
         6 ProGluValProGluLeuPheArgGluValLys............. 17

        83 TGAGGGCGCCCCGCGGGTCTGGTCTGGGCCGGGGAACTCCGGGACGGCGGC 132
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        17 17

       133 CGCAGCAGGGCTACGGATCTCTCGGCCCCCTTGCCCTTCAAATCACCTCTAG 182
            ::: ||||||||||::::::: |||
        18 ..TyrTyrAlaValGlyTyrAspIleAspPro................ 26

       183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAAGCGGAGGAAGTTTCCTA 232
            |||||||||||||||||||||||||||||||||||||||||||||||||
        27 ...GlnValIleGlnLeuLeuLysAlaGlyLysAlaLysLysGluValSerTyr 42

       233 CAATGCACTAGCCTTCACACATAATCTCAGAGGATGGGGACAATCCAGAGG 282
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        42 rAsnAlaLeuAlaSerHisIleIleSerGluAspGlyAspAsnProGluVal 59

       283 TGGGAGAAGCTCGGGAAGTCTTTGACTTACCTTGTTGTAAAGCCTTCTTGG 332
            ||||||||||||||||||||||||||||||||||||||||||||||||||
        59 aIGlyGluValArgGluValPheAspLeuProValVallLysProSerTrp 75

       333 GTCGATTCTGTCCGTTCAGTGTGGAACTCTTCTGCCAGTAAATGGTTTTTC 382
            ||| |||||||||||::||||||||||||||::::::||||||||||

```
 76 ValThrLeuSerValGlnCysGlyAlaLeuLeuProValAsnGlyPheSe  92
383 TCCAGAATCACAATGTCAGAATTTTTTGGAATCACTGCCTTCTCAGG      432
    |||||||||||||||||||||||||||||||||||||||||||||
 92 rProGluSerCysGlnIlePhePheGlyLeuThrAlaCysLeuSerGln.  108
109 ..............................................      109
433 GTGTTGGATACAAGCTGGAGCGTCTTTGTTGTGGGAGTCTTCCCA       482
    |||
109 Val..                                               109
483 AGGAGAGGTAGGGAGGAGGCTTGTTCCAGCAGAAGTTGGGAACCACAGAG  532
109 ..............................................      109
533 ATCATCTGCCTTCTTCTGACCCGGTTATTGATGAGGCCTCTGTT        582
109 ..............................................      109
583 GTAAGTGCTGGGTGTGTCATCTGAGACAGAGAAGTGCCCTTGGGCTTTGGT 632
    |||||||||||||||||||||
110  ?    ......SerSerGluAspArgSerAlaLeuTrpAlaLeuVa    121
633 TACGTTCTTAAGGGGGGAGAGATTGCCAGCTAACCCTCAATAAGAAATGCACGC 682
    ||||||||||||||||||||||||||||||||||||||||
121 lThrPheHisGlyGlySerCysGlnIleLeuAsnLeuAsnLysLysCysThrH 138
683 ATTTGATTGTTCCAGAGCCAAAGGGGGGAAAATACGAATGTGCTTTAAAG 732
    |||||||
138 isLeuIleValProGluProGlyLysGlyLysTyrGluArgAlaValVal 154
733 CGAGCAAGTATTAAAAATTGTGACTCCTGACTGGGTTCTGAATTGCGTATC 782
    ||||||||||||||||||||||||||||||||||||||||||
155 ArgThrSerIleLeuLysIleValThrProAspTrpValLeuAspCysValSe 171
783 AGGAGAAAACCAAAAAGACGAAGCANTTAATCATCCTCGTCTGANTATTT 832
    |||||| ::::||||||||||||||||||||||||||||||||||||||
171 rGluLysSerArgArgLysAspGluAlaPheTyrHisProArgLeuIleLeT 188
833 ATGAAGAGGAGGAGAGAGAGGAAGAGAGGAGGAAGTAGAAAATGAG       882
    |||||||
188 yrGluGluGluGluGluGluGluGluGluGly........AspAsnGlu    201
883 GAACAAGATTCTCAGAAGAGGGTAGTACAGAGAGAAGTCAAGCCCTGC    932
    |||||||||||||||||||||||||||||||||||||||||
202 GluGlnAspSerGlnAsnGlyGlySerThr...    GluLysSerSerValAl 217
933 CAGCTCCTCAAGAACAGGGTCTCCCTTCCAGGTGACCAGCAGTTTCACCTAAAT 982
    ||||||||||::::||||||||||
217 aSerSerAlaValAlaSerProAla...GluGlnProCysSerProLysSP   233
983 CCAACACTGAAAAAAATCTAAAGGGGAATTAAATGTTTGAATTCTTCAGAT 1032
    ||||||||
233 roArgAlaGluValSerLysGlyGluLeuMetPheAspAspSerSerAsp 249
1033 TCATCACCGGAAAAAACAGAGAGAAAATTTAAAGGTGACCCGGCCGAAGT 1082
    |||||||||||||||||||||||||||||::||||||::
250 SerSerProGlyGlyGluArgSerLeuAsnTrpAlaProAlaGluVal    266
1083 CCCACAGTTAGCTCGCAAAAAACCGCAGGCGTGCCTCAGGGAAAGGAGCCTG 1132
    ::|||                |||||||||||||||||||||||||
266 aProProLeuAsnThrAlaGlnArgAArgLeuProGlnGlyTyrLysGlyProG 283
1133 GGTTGATTAACTTGTGTGCCAAATGTCCACCCGTCCAGGTAACATTTTG 1182
    |||||||||||::|||||||||||||
283 lyLeuIleAsnLeuCysSerAlaAsnValProProValProGlyAspProLeu 299
1183 CCCCCTGAGGTCCCGGGGTAATTTATGGGCTGTGTTGGACAAAACCTCCAAG 1232
    ||||||::::::|||||||||::||||
300 ProProAspMetArgGlyAsnLeuMetAlaProGlyAsnLeuGlnAs      316
  3 TTCTGAAAGANTCAGAAANGAATAGCTACCCTGGACTCCAGCGCTGTAGGACAC 1282
    ||||||||||||::::|||||||||||||||
333 LuArgSerGluIleLeuGlyTyrThrTrpSerProAlaValArgGlnThrT    333
```

```
1283 TGAGGAATATTACTAATAATGCTGACAATTCAGCAGATGAAACCGGCCATCA 1332
     |||||||||||||||||||||||::|||||||||||||||::||||||||
 333 euArgAAsnIleThrAsnAsnAlaAspIleGlnIleIleAsnArgProSer  349
1333 AATGTAGCACAATAATCTTTACAGACTCTTTCAGCCACCTACAGAAAAATTTTAGA 1382
     |||||||||||||||||||||||::||||||||||||::::||||
 350 AsnValAlaHisIleLeuGlnSerLeuSerAlaProThrLysSerLeuG1   366
1383 ACAGCAGGTGAATCACAGCCCAGCAGGGACTACAAATGCCAAATGCAGTGC 1432
     ||||||||||||||||||||||||||||||||||::||||
 366 uGlnGlnValAlaAArgGlyGlnGlyHisProAsnAlaSerAlaVall.L  383
1433 TGTTTAGCCAAGTGAAAAGTGACTCCAGAGACAGACATCAGTACAGCAG 1482
     ::|||||::||::::|||            ::||||||||||::
 383 euPheGlyGlnAlaLysGlyAlaProGluThrHisValLeuGlnHis    399
1483 CAGCAGCCCCAGCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCC 1532
     :::|                    ||||||||||||||||||||||||
 400 lnGlnGlnProGlnGlnGlnGlnGlnGlnGlnHisProProSerHisGlnPr   416
1533 CCAGCAGAATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGAATCTCTCAG. 1581
     |||||||||||||||||||||||::||||||            :::||
 416 o...GlnIleMetGlnLeuGlnLnGlnGlnGlnGlnGlnGlnGlnGlnG   432
1582 CAACCTTACCCCCAGCAGCCCGCGCCGCCATCCAATTTCA...         1617
     |||||||||||||||||||||||||||||||||||
 449 GlnValHisGlnHisGlnPheSerGlnGlnGlnGlnProGlnGl       465
1617 ..............                                     1617
 449 ...GlnIleMetGlnLeuGlnLnGlnGlnGlnGlnGlnGlnGlnGl      465
1618 CAGCAGCAGCAGCAGCAGCAGCAGCAG...                     1641
 515 nGlnHisAsnLeuLeuLeuGlnGlnGlnGlnLeuGlnArgL          532
1641 ..............                                     1641
 532 euGlnGlnGlnGlnGlnMetGlnAsnGlnAlaAlaHisLeuSerGlnAla 548
1642 CCACCACCATGGCCTCAGCAGCAATCAGCTTTTTGGACATG          1681
     |||||               ||||||||||||||||||
 549 SerGlnAlaLeuGlnHisGlnValLeuProGlnLeuGlnLeuProLeuGlnLeuSe 565
1682 ATCCACGCAGTGGACAGATTCCAGAGAAGAGGCTTCTTATTGGGATGTGTGTTT 1731
     |||::||||||||||||||||::|||||||||||||||||||
 565 rLeuGlnLeuProGlnProGlnLeuGlnGlnLeuGlnLeuPheGlyHisA    582
1732 GCAATTCCGGATTATACCAGAGACGATGTTCTGAATAAGCAAACTGCTGGGCCAC 1781
     ||||||||||||||||||||::||||||||||||||||||||||||
 599 AlaIleAlaIleAspTyrTyrProGlnMetSerAspProLysGlnLeuLeuAlaTh 615
1782 CTGGAAAAAGAGAATAATCAGGGACACAAGGGGCACTGTTGACCCCACCTTCA 1831
     ||||||||||||||::||||||||||||||||||||||||||||||
 615 rlTrpLysArgIleIleGlnIleAlaHisSerGlyGlyThrValAspProThrPheT 632
```

```
1832 CGAGTCGGATGCCACGCCACCTTCTCTGAGAGTCAAGTCAAGTCAGCGCGTAT 1881
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 632 hrSerArgCysThrHisIleuLeuCysValAlaSerGlnValSerSerMetTyr 648

1882 GCACAGGCANATAAGAGAAAGAGAAGAGATGTTACTGACACTGGTTAAA 1931
     ::|||||||||||||||||||||||::|||::||||||||||:||:|||||||
 649 ThrGlnAlaLeuAsnArgGluArgLysArgCysValThrAlaHisTrpLeuAs 665

1932 CACAGTCTTAAAAAAAAAAAAATGGTACCCGCGCCACCCGAGCCCTTCACT 1981
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 665 nIhrValLeuLysLysLysLysLysLeuMetProProHisArgAlaLeuHisP 682

1982 TCCCAGTGGCGCTTCCCACCCAGGAGGAAAGCCANTGTTCACAGCATATTATT 2031
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 682 heProValAlaLeuProProGlyGlyLysProCysSerGlnHisIleIle 698

2032 TCTGTGACTGGGAGTTGTTGAATGACAGAGATGACCTAAAATTAAATGGC 2081
     |||||||||::||||||||||||||||||||||||||||||||||||||||||
 699 SerValThrGlyPheValAspAspArgAspAspLeuLysLeuAsnMetTAl 715

2082 TTAATTGGCAGGTGCCAAATATACGGGGTTATCTATGCCCGACGAACACAG 2131
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 715 aTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrV 732

2132 TCCTCANTCTGTAAAGAACCAACTGGTTTAAAAGTATGAAAAAGCCAAAGAG 2181
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 732 aLeuIleLysLysGluProSerGlyLeuLysTyrGluLysAlaLysGlu 748

2182 TGGAGGATACCCTGTGTCAACGCCCACCAGTGGGCTTGGGCACATTCTTCTGGG 2231
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
 749 TrpArgIleProCysValAsnAlaGlnHisTrpLeuGlyLysAspIleLeuLeuGl 765

2232 AAACTTTGAGGCACTCGAGGCGATTCAGTCAGTAGTCGCTACACGGCATTCA 2281
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 765 yAsnPheGluAlaLeuGluArgIleGlnTyrSerArgTyrThrAlaPheA 782

2282 GTCTGCAGGATCCATTGCCCCTACCCAGCACATTGACTHTAAATCTTTTA 2331
     ::|||  ||||||||||||||||||||||||||||||||||||||||||||
 782 snMetProAspProPheValProThrProHisLeuValLeuGlyTyrLeuLeu 798

2332 GATGCGTTTGAGAGAGTTCCCTTTAAAAGTGTCTCAGAGTGTTGATGAGTAT 2381
     |||||||||||||||||||||::|||||||||||||||::||||||||:::
 799 AspAlaIatTpArgThrProLeuLysValThrAlaGluLeuLeuMetGlyVa 815

2382 AAGACTACCTTCCCAAAACTGAAACAAGAATAGAAGTAGCCTAAATGTCCAGCCTT 2431
     :||:|||    |||||||||||||||||||||||||||||||:::|||:|||
 815 lArgLeu***ProLysLeuLysProAsnGluValAlaAsnIleGlnLeuS 832

2432 CTTCCCAAAGAGCCAGAATTGAAGACGTACCACCCCCCCACTAAAAGCTA 2481
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 832 erSerLysArgAlaArgIleGluAspLeuProProThrLysLysLeu 848

2482 ACTCCAGAATTGACCCCTTTTTGTGCTTTTCACTGGATTCGACCCTGTCCA 2531
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 849 ThrProGluLeuThrProLeuValLeuPheHisArgIlePheGluProValGl 865

2532 GGTTCAACAGTATAATAAGGAGCTCTACATTCTTGGTGGAGGAGGTTGCGG 2581
     |||||||||||||||||||||||||||||||||||||||||||||||||||
 865 nValGlnGlnTyrIleLysGluLeuTyrIleLeuGlyGlyGlyValAlaG 882

2582 AGTCTGCACAGAAGAGTGCACACCTCATTGCCAGCAAAGTGACTCGCACC 2631
     ||  :||::|||||||||||||||||||||||||||::|||||:||||||
 882 luCysThrLysLysCysThrHisIleuLeuAlaSerLysValThrArgThr 898

2632 GTGAAGTTCCTCGACGGCGATTCGTCGTGAAGCACACTAGTGACCGCCAGA 2681
     |||||||||||||||||||||::|||||||||||||||||::|||:||::::
 899 ValLysPheLeuThrAlaIleSerValAlaLysHisIleValThrProAs 915

2682 GTGGCTGGAAGAATGCTTCAGGTGTCAGAAGGTTCATGAGAGCAGAACT 2731
     :|||||||||||||||||||||||::|||||||||::|||::|||||::::
 915 pTrpLeuGluLysPheLysArgGlnThrPheIleAspGluGlnAsn 932

2732 ACANTCTCCGAGATGCTGAGGCAGAAGTACTTTCTCTTTCAGCTTGGAA 2781
     |||||||||||||||||||||||||||||||||||||||||||||||||||
```

seq_name: sp_human:O15404

```
 932 YrIleLeuArgAspAlaGluValAlaGluValAlaLeuPheSerPheSerLeuGlu 948
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
2782 GAATCCTTAAAACGGGCGACACGTTCTCCACTCTTTTAAGCAAAATATTT 2831
     |||||||||||||||||||||||||||||||||::|||||||||||||||||
 949 GluSerLeuLysArgAlaHisValSerProPheHisLeuSerThrLeuYrrPh 965
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
2832 TTACATCACCTGGAAATCTGCCCCAAGTCTTTCCACTATGAAGGCAATCG 2881
     |||||||::||||||||||||||||||||::||:|||||||||||||||||
 965 eTyrIleThrProGlyTrpIleCysProSerLeuAlaThrMetLysAlaIleV 982
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2882 TAGACTGTGCGAGAGAAAGTGTTATCCAGCACCAGCCACTTCCCGGAAG 2931
     |||||||||||||||||||||||||||||||||::||||||::||||||||
 982 alGlyCysAlaGlyLysValLeuAlaLysGlySrProSerPheArgLys 998
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2932 CTCATGGACACAAGCACAGAACTCGAGTTGTCGGGAAAATAATTTAATAATC 2981
     |||||||||||||||::|||||||||||||||||||||||||||||||||||
 999 LeuMetGluHisLysGlnAsnLysSerLeuSerGluIleLeuIleSe 1015
     |||||||||||||||||||||||||||||||||||||||||||||||||||
2982 CTGTGAAAATGACCTTCATTTATGCCGAGAATATTTGGCGAGGGCATAG 3031
     ||||||||||||||||||||||||||||||||||||||||||||||||||
1015 rCysGluAsnAspLeuHisIleLeuCysArgGluTyrPheIleAlaArgGlyIleA 1032
     |||||||||||||||||||||||||||||||||||||||||||||||||||
3032 ATGTTCACAATCCAGAGTTCGTGTTCTGACTGGAGTGCTCCTCACTCAAAGCTG 3081
     |||||||||||||||||||||||||||||||::|||||||||||||||||||
1032 spValHisAsnAlaGluProPheValLeuuThrGlyValLeuuThrGlnThrLeu 1048
     |||||||||||||||||||||||||||||||||||||||||||||||||||
3082 GACTANGAAATCANTATAAGTTAAC 3105
     |||||||||||||||||||||||||||
1049 AspTyrGluSerTyrLysPheAsn 1056
```

```
 874 GAAAATGAGGAACAGAGATTCTCAGAAGAGGGTAGTACAGATGAGAAGTC 923
     |||||||||||||||||||||||||||||||||||||||||||||||||||
```

seq_documentation_block:
```
ID   Q90WJ3               PRELIMINARY;      PRT;   1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE 1256 AA; 144187 MW; E7795I2C1A42DE1 CRC64;
```

seq_name: sp_vertebrate:Q90WJ3

alignment_scores:
```
        Quality:  3425.50           Length:  1332
        Ratio:    4.112             Gaps:    20
Percent Similarity: 62.538    Percent Identity: 53.829
```

alignment_block:
US-09-664-641-10 x Q90WJ3

```
Align seq 1/1 to: Q90WJ3         from: 1    to: 1256

165  CCTTCAAAATCACCTCTAGGCACAGGTTATTCAGCTCTCCAAGGCTGGAAA   214
     |||  |:|:  ||   |   |:|:|: ||||||||||:|||||||||||||
20   ProValGlyAspLeuAspProLysValValGlnMetLeuLysAlaGlyLly  36
```

OM of: US-09-664-641-10 to: SPTREMBL_19.*   out_format : pfs

Date: Jun 4, 2002 5:49 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL-frame+.n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09664641/runat_04062002_110933_9250/app_query.fasta_1.3697
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09664641.#CGN1_1_312 -NCPU=6 -ICP=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19.*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
sp_rodent:Q920W6	+	4059.50	5875.12	0.0	1056	!	Q920w6 mus musculus (mouse).
sp_human:O15404	+	3841.00	5561.10	0.0	744	!	O15404 homo sapiens (human). ca
sp_vertebrate:Q90WJ3	+	3425.50	4951.40	1.5e-268	1256	!	Q90wj3 xenopus laevis (african
sp_human:Q96HP2	+	2039.00	2947.67	1.9e-156	391	!	Q96hp2 homo sapiens (human). un
sp_invertebrate:Q9VUB6	+	852.00	1205.56	4.6e-60	1798	!	Q9vub6 drosophila melanogast
sp_invertebrate:Q94046	+	524.00	734.07	1.4e-33	1074	!	Q94046 caenorhabditis elegan
sp_invertebrate:Q9U370	+	524.00	734.05	1.4e-33	1076	!	Q9u370 caenorhabditis elegan
sp_human:Q14676	+	342.50	463.12	8.9e-19	2089	!	O14676 homo sapiens (human). k
sp_human:Q96QC2	+	340.50	460.21	1.3e-18	2090	!	Q96qc2 homo sapiens (human). k
sp_plant:Q9LVF7	+	316.50	432.68	8.9e-17	1041	!	Q9lvf7 arabidopsis thaliana (n
sp_rodent:Q91YX3	+	283.50	382.38	4.5e-14	1296	!	Q91yx3 mus musculus (mouse). s
sp_plant:Q92R18	+	265.50	361.78	1.1e-12	765	!	Q9zr18 arabidopsis thaliana (mc
sp_human:Q9UEB9	+	263.50	352.22	2.0e-12	1435	!	Q9ueb9 homo sapiens (human).
sp_human:Q92547	+	263.50	351.40	2.0e-12	1550	!	Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5	+	229.00	307.74	1.0e-09	838	!	Q9p5j5 neurospora crassa. cons
sp_invertebrate:Q9VFR7	+	225.50	303.46	1.9e-09	776	!	Q9vfr7 drosophila melanogaste
sp_invertebrate:Q24163	+	214.00	291.75	5.5e-09	483	!	Q24163 drosophila melanogast
sp_invertebrate:Q9VED3	+	220.50	291.76	5.6e-09	1180	!	Q9ved3 drosophila melanogaste
sp_invertebrate:O44011	+	219.50	288.00	7.2e-09	1457	!	O44011 dictyostelium discoid
sp_invertebrate:Q9W4E8	+	216.00	293.94	9.6e-09	517	!	Q9w4e8 drosophila melanogaste
sp_plant:Q93VQ3	+	215.50	283.70	1.4e-08	1272	!	Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853	+	214.00	293.23	1.3e-08	420	!	O76853 dictyostelium discoide
sp_fungi:P79025	+	214.00	291.75	1.4e-08	483	!	P79025 emericella nidulans (asp
sp_virus:Q98148	+	214.00	282.47	1.9e-08	1162	!	Q98148 kaposi's sarcoma-associ
sp_vertebrate:Q9VP17	+	213.50	274.60	2.6e-08	2285	!	Q9vp17 drosophila melanogaste
sp_vertebrate:Q9YI2	+	212.50	275.54	2.9e-08	1822	!	Q9yi2 fugu rubripes (japanese
sp_invertebrate:Q9YOC9	+	212.00	283.02	2.4e-08	838	!	Q9yoc9 dictyostelium discoide
sp_invertebrate:O18346	+	210.50	278.87	3.4e-08	1010	!	O18346 drosophila hydei (fru
sp_virus:Q9QR71	+	210.00	276.96	3.9e-08	1129	!	Q9qr71 kaposi's sarcoma-associ
sp_virus:Q9VMS2	+	208.50	266.64	6.7e-08	2439	!	Q9vms2 drosophila melanogast
sp_invertebrate:Q95ZG5	+	207.00	262.29	9.6e-08	2994	!	Q95zg5 drosophila melanogast
sp_invertebrate:Q9V7C5	+	206.50	273.98	6.9e-08	925	!	Q9v7c5 drosophila melanogaste
sp_invertebrate:Q95M8	+	206.50	268.20	8.4e-08	1598	!	Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641	-	205.50	288.88	4.8e-08	197	!	Q17641 caenorhabditis elegans
sp_invertebrate:Q9VN59	-	205.00	271.70	9.2e-08	934	!	Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485	-	204.50	274.64	8.9e-08	660	!	Q94485 dictyostelium discoid
sp_invertebrate:Q9N8T2	-	203.00	255.01	2.1e-07	3439	!	Q9n8t2 trypanosoma brucei. h
sp_human:O15410	+	202.50	271.86	1.3e-07	652	!	O15410 homo sapiens (human). ca
sp_human:O75557	+	202.50	259.89	1.9e-07	2023	!	O75557 homo sapiens (human). o
sp_human:Q9UND7	+	202.50	259.89	1.9e-07	2023	!	Q9und7 homo sapiens (human). o

alignment_scores:
          Quality: 4059.50   Length: 1142
          Ratio: 4.572   Gaps: 11
          Percent Similarity: 77.758   Percent Identity: 70.928

alignment_block:
US-09-664-641-10 x Q920W6

Align seg 1/1   to: Q920W6   from: 1   to: 1056

```
33   CGGGAATTCCCGGTGTCGACGATTTCGTCGGAGTGGCAAGAGACGGGGCGGGCCTGC   82
     ||||||  || ||||| :|| ||||||| || | ||||:||:::
6    ProGluValProGluGluLeuPheArgGluValLys...............   17

83   TGAGGGCGCGCCCCGGCGGGGTTCGTGGTCTGGGCCCGGGGGAACTCCGGGACGGCGC  132
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
18   .TyrGlyTyrAlaValGlyAspIleAspPro...............  26

133  CGCAGCAGGCGCTACGGATCTCTCGGGCCCCGCCCTCGCCCTTCAAATCACCTCTAG  182
     |||| ||| :|| |||||| :::| | ||||||||||||||||||||||||||
27   ...GlnValIleGluIleLeuLeuLysAlaLysValLysGluValSerTyr  42

183  GCACAGGTTATTCAGTCTTCTCAAGGCTGGAAAAGGGAAGGGAAGGTTTCCTA  232
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
42   rAsnAlaLeuAlaSerIleIleIleSerGluAspGlyAspAsnProGluV  59

233  CAATGCACTAGCCCTCACACAATAACTCAGAGGATGGGGACGACAAGCCAGAGG  282
     |||||||||||||||||||||||||||||||||||||||||||||||||||||
59   aGlyHisGlnLeuGluLeuGlySerGluGluAspGlyAspLysAlaGluThr  75

283  TGGGAGAAGCTCGGGAAGTCTTTGACTTACCTGTTGTTAAAGCCTTCTTGG  332
     |||||||||||||||||||||||||||||||||||||||||||||||||||
75   alGlyGluAlaArgGluValPheAspGluProValAlaValLysProSerTrp  75

333  GTGGATTCTGTCCCGTTCAGTGTGGGAACTCTTCTGCCCAGTAAATGGTTTTTC  382
     |||    |||||||||||||||||||||||||||||||||||||||||||||||||
     ||||||||||||||||||||||||||||||||||||||||||||||
```

```
  76 ValThrLeuSerValGlnCysGlyAlaLeuLeuProValAsnGlyPheSer  92
 383 TCCAGAATCAAGTCAGATTTTTTTGGAAACACTGCTGCCTTCTCAGG 432
  92 rProGluSerCysGlnIlePhePheGlyLeuThrAlaCysLeuSerGln. 108
 433 GTGTTGATACAAGCTGGAGCTCTTGTGTTGTTGGAGTCTTCAGAGCTCTCCCA 482
 109 Val.............................................. 109
 483 AGGAGGAGGTAGGGAGGGAGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAG 532
 109 .............................................. 109
 533 ATCAATCTGCCTTCTTCTGACCCGGTATTGATGCAGGCTTCTTGTT 582
 109 .............................................. 109
 583 GTAAATGTGCTGGGTGTCATCTCAAGACAGAAGTGCCCTGTGTGGGCTTTGGT 632
 110 ?     .     SerSerGluAspArgSerAlaLeuThrProAlaLeuVa 121
 633 TACGTTCTATGGGGGGAGAATTGCCAGCTAACCCTCAATAAGAAATGCACGC 682
 121 lThrPheHisGlyGlySerCysGlnLeuAsnLeuAsnLysLysCysThrH 138
 683 ATTTGAATTGGTTCCAGAGCCAAAAGGGGGGAAAANCGAATGTGCTTTAAAG 732
 138 isLeuIleValProGluProLysGlyCysLeuLysThrValGluArgAlaVaLLys 154
 733 CGAGCAAGTATTATTAAAATTGTGACTCTCTGACTGGGTTGTTGGGAATTGGGTATC 782
 155 ArgThrSerIleLysIleValThrProAspTrpValLeuAspCysValSe 171
 783 AGAGAAAACCAAAAAGGAGGAAGCAATTTTTTAATCANTCCTGCTCGAATTATTT 832
 171 rGlyLysArgArgLysAspGluAlaPheTyrHisProArgLeuIleLeu 188
 833 ATGAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGTAGAAAATGAG 882
 188 yrGluGluGluGluGluGluGluGly.......AspAsnGlu 201
 883 GAACAAGATTCTCAGAATGAGAGGGTAGTACAGATGAAGAGTCAAGCCCTGC 932
 202 GluGlnAspSerGlnAsnGluGlySerThr...GluLysSerValAl 217
 933 CAGCTCCTCAAGAAGGGTCCCCTTCCAGGTGACCCCACAGTTTTCACCTAAAT 982
 217 aSerSerAlaValAlaSerProAla...GluGlnProCysSerProLysP 233
 983 CCAACACTGAAAAATCTAAAGGGGAATTAAANGTTTGATGAATTCTTCAGAT 1032
 233 roArgAlaGlyValSerLysGlyGluGluMetPheAspAspSerSerAsp 249
1033 TCATCACCGGAAAAACGGAGAGAAATTTAAAACTGGACCCCGGCCGAAGT 1082
 250 SerSerProGluLysGlnGluArgSerLeuAsnTrpAlaProAlaGluAl 266
1083 CCCACAGTTAGCTGCAAAAACGCAGGCTGCCTCAGGGAAGAAGAGCCTG 1132
 266 a ... ProProLeuAsnThrAlaGlnArgLeuProGlnGlyLysGlyPro G 283
1133 GGTTGAATTAACTGTGTGCCCAATGTCCACCCGTTCCCAGGTAACATTTTTG 1182
 283 lyLeuIleAsnLeuCysSerAlaAsnValProValProGlyAspIleLeu 299
1183 CCCCCTGAGGTCCGGGTAATTTAATGGCTGCTGCTGACAAAACCTCCAAG 1232
 300 ProProAspMetArgGlyAsnLeuMetAlaAlaProGlyGlnAsnLeuGlnAs 316
 ??? TTCTGAAAAGTATCAGAAAAATGATAACTACCTGGAGTTCCAGCTGACTCGTACGGACAC 1282
     ...GlnArgSerGlu...IleLeuGlyThrThrTrpSerProAlaValAlaArgThrT 333
```

```
1283 TGAGGAAATTATTACTTAATAATGCTGACATTGCAGAGGAGAGACCGGCCATCA 1332
 333 euArgArgAsnIleThrAsnAsnAlaAspIleGlnIleIleAsnArgProSer 349
1333 AATGTGACATATCTTACAGACTCTTTCACCACCTACGAAAAATTAGA 1382
 350 AsnValAlaHisIleIleGlnSerLeuSerAlaProThrSerSerLeuGl 366
1383 ACAGCCAGGTGAATCACACCAGCAGGACAATACAAAATGCCAAATCCAGTGC 1432
 366 uGlnGlnValAlaArgGlyGlnGlyHisProAsnAlaSerAlaValAll 383
1433 TGTTTAGCCAAGTGAAAGTGACTCCAGCAGCAGCAGCACACATGCTACAGCAGCAG 1482
 383 euPheGlyGlnAlaLysGlyAlaProGluThrHisValLeuGlnGlnHis 399
1483 CAGGCAGGCGCCCAGCAGCAGCAGCAGCACCACCCGTTTACACCTTCAGCC 1532
 400 HisProProGlnGlnProGlnGlnHisProAlaLeuHisLeuGlnPr 416
1533 CCAGCAGATAATTGCAGCTCCACAGCAGCAGCAGCAGCACAGCAGCAGCTCTCAG 1581
 416 o...GlnIleMetGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnHisGlnG 432
1582 .....CAACCTTACCCCACCAGCAGCGCCGCGCCCATCCAATTTCA... 1617
 432 lnGlnGlnProGlnProTyrProGlnProSerHisGlnPhePheGlnGln 448
1617 ........................................ 1617
 449 GlnValHisGlnHisGlnPheSerGlnGlnLeuGlnLeuGlnPheProGlnGl 465
1617 ........................................ 1617
 465 nProLeuHisProGlnGlnLeuHisArgProGlnGlnGlnLeuGlnP 482
1617 ........................................ 1617
 482 roPheGlnGlnGlnHisGlnLeuGlnGlnHisGlnGlnGlnGln 498
1617 ........................................ 1617
 499 GlnGlnLeuGlnHisHisGlnLeuAlaGlnLeuGlnGlnGlnGlnGlnGl 515
1618 .....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAACAGCAG..... 1641
 515 nGlnHisAsnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnArgL 532
1641 ........................................ 1641
 532 euGlnGlnGlnGlnGlnMetGlnAsnGlnAlaAlaHisLeuSerGlnAla 548
1641 ........................................ 1641
 549 SerGlnAlaLeuGlnHisGlnValLeuProGlnGlnProLeuGlnLeuSe 565
1642 .....CCACCACCATCGCCTCAGCAGCANCACACTTTTTGGACANG 1681
 565 rLeuGlnProProGlnGlnGlnGlnLeuGlnLeuHisGlyHisSe 582
1682 ATCCAGCAGCTGGAGATTCCAGAAGAGAAGGCTTCTAATGGAGTGTGTTT 1731
 582 spProAlaValGlyIleLeuProGlnLeuSerPheLeuLeuGlyCysValPhe 598
1732 GCAATTGCGGATTATCCAGAGACGAGTCTGGATAAGCAACNCTGCTGGGCCAC 1781
 599 AlaIleAlaAspThrYrProGluGlnMetSerAspLysGlnLeuLeuAlah 615
1782 CTTGGAAAAGGATAAANTCCAGGCACACATGGCCGCGCACTGTTGACCCCACCTTCA 1831
 615 rTrpLysArgIleIleGlnIleHisAlaHisGlyGlyThrValAspProThrPhe 632
```

```
1832 CGAGTCGGATGCACGCCACCTTCTCTGTTGAGGAGTCAAGTCGCAGCCGTAT 1881
          |||||||||||||||||||||||||||||||||||||||||||||  |||
 632 hrSerArgCysThrHisLeuLeuCysAlaSerGlnValSerSerMetTyr   648

1882 GCACAGGCAAATAAGAGAAAGAAAAGAGAATGTTGTTACTCCACTGGTTAAA 1931
          :::|||||||:|||||||||||||||||||||||||||:::|||:||||
 649 ThrGlnAlaLeuArgGluArgLysArgCysValThrAlaHisTrpLeuAs   665

1932 CACAGTCTTAAAAAAAAAAAAAAATGGTACCCGCACCAGCGCCCTTCACT 1981
          ||||||||||||||||||||||:|||||||||||:::||||||||||
 665 nThrValLeuLysLysLysLysLeuMetProProHisArgAlaLeuHisP   682

1982 TCCCAGTGGCCTTCCACCCAGGAGGAGAAAGCCATGTTCACACGACATATTATT 2031
          ||||||||||||||||||||||||||||||||||||||||||||||||
 682 heProValAlaPheProProGlyGlyLysProCysSerGlnHisIleIle   698

2032 TCTCGTGACTGGATTTGTTGATGAGTGACACAGAGATGACCTAAAATTAATGGC 2081
          ||||||||||||||||||||||||||||||||||||||||||||||
 699 SerValThrGlyPheValAspAspAsnAspArgAspAspProLeuLysMetI   715

2082 TTATTTGGCAGGTGCCAAATATACGGGTTATTAGTGGGTTGYLeuCysArgSerAsnThrV 2131
          ||||||||||||||||||||||||||||||||||||||||||||||
 715 aTyrLeuAlaGlyValAlaLysTyrThrGlyTyrLeuCysArgSerAsnThrV 732

2132 TCCTCATCTGTTAAAGAACCACTGGTTTAAAGTATGAAAAAAGCCAAAGAG 2181
          ||||||||||||||||||||:::|||||||||||||||||||||
 732 aLeuLeuIleCysLysGluProSerGlyLeuLysTyrGlyLysAlaLysGlu 748

2182 TGGGAGGATACCCTGTGTCAAAGCCCAGTGGCTTGGGACAATTCTTCTGGG 2231
          ||||||||||||||||||||||||||||||||||||||||||||||
 749 TrpArgIleProCysValAsnAlaGlnTrpLeuValAspIleLeuLeuLeu1 765

2232 AAACTTTGAGGCACTGAGGCACGAATTCAGTATAAGTCGGCTACACGGCATTCA 2281
          |||||||||||||||||||||||||:||||||||||||||||||||:
 765 yAsnPheGluAlaLeuLeuArgGlnValGlnTyrSerArgTyrThrAlaPheA 782

2282 GTCTGCAGGATCCATTGGCCCTACCCCAGCGATTTAGTTTTTAAAATGTTTTA 2331
          ||||||||||||||||||||||||||||||:::|||||||:||||||
 782 snMetProAspProPheValProThrProHisLeuValLeuGlyLeuLeu   798

2332 GATGCTTGGGAGAGTTCCCTTAAAAGTGTCTCGCAGAGTTGTTGAICGAGTAT 2381
          |||||||||||||||||||||||||||||||||||||||||||||||
 799 AspAlaTrpGluArgThrProValLysValThrArgGluLeuLeuMetGlyVa 815

2382 AAGACTACCTCCCCAAACTGAAACAGAATGAAGTAGCTAAATGTCCAGCCTT 2431
          ::||| |M|   |||||||||||||||||||||||||:||||:  |
 815 lArgLeu***ProLysLeuLysSerProAsnGluValAlaAsnIleGlnLeuS 832

2432 CTTCCAAAAGAGCCAGAATGAAGAGCGTACCACCTCCCACTAAAAAGCTA 2481
          ||||:||||||||||||||||||||||||||||:::||||||||||
 832 erSerLysArgGlnAsnGluGluArgIleGluLysProProThrLysLysLeu 848

2482 ACTCCAGAATTGACCCCTTTTGTGCTTTTCACCTGGATTCGAGCCTGTCCA 2531
          |||||||||||||||||||||||||||||||||||||||||||||||
 849 ThrProGluLeuThrProLeuValLeuPheThrGlyPheGluProValGl   865

2532 GGTTCAACAGTATATTAATTAAGAAGCTCTACATTCTTGGGTGGAGAGGTTGCGG 2581
          ||||||||||||||||||||||||||||||||||::||||||||||||
 865 nValGlnGlnTyrIleLeuLysLysLeuTyrIleLeuGlyValGlyValAlaG 882

2582 AGTCTGCACAGAAGTGCACACACCTCATTGCCCAACAAAGTGACTGGCACC 2631
          :::||||||||||||||||||||||||||||||||||||||||||||
 882 luCysThrLysLysCysThrHisLeuIleAlaSerLysValThrArgThr   898

2632 GTGAAGTTCCTGACGGCGGATTTCGTCGTGAAGCACATAGTGACGGCCAGA 2681
          |||||||||||||||||||||||||||||||||||||||||||||||
 899 ValLysPheLeuThrAlaIleSerValValLysHisIleValThrProHis 915

2682 GTGGCCTGGGAAGAAAATGCTTCAGGTGTCAGAAGTGTCATTGATGAGGAACT 2731
          :|||||||||||||||||||||||||||||||||||||||||||
 915 pTrpLeuGluGluAsnCysPheSerLysArgGlnHisPheIleAspGluGlnAsnT 932

2732 ACAATCTCCGAGAATGCTGAGGCGCAGAGAAGTACTTTCTCTTTCAGCCTTGGAA 2781
          ||||||||||||||||||||||||||||||||||||||||||||||
```

```
 932 yrIleLeuArgAspAlaGluAlaGlnGluValLeuPheSerPheSerGluLeu 948
          |||||||||||||||||||||||||||||||||||||||||||||
2782 GAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAAAGGCCAAAATATTT 2831
          |||||||:::||||||||||||||||:::||||:||||||||||||
 949 GluSerLeuLysArgAlaHisValSerProPheHisLysThrLysTyrPh 965
2832 TTACATCACACCTCGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCG 2881
          ||||||||||||||||||||||||||||:::|||||||||||||
 965 eTyrIleThrProGlyIleCysProSerLeuAlaThrMetLysAlaIleV 982
2882 TAGAGTGTGCAGGAGGAAAGGTGTATCCAAGCAGCCCATCTTTCGGAAG 2931
          |||||||||||||||||||||||||||||||||||||||||||||
 982 alGluCysAlaGlyGlyLysValLeuAlaLysGlnProSerPheArgLys 998
2932 CTCATGAGCACAGCACAAACTCGACTTTTGTCTCGAGAATAATTTTAAATAAC 2981
          |||||||||||||||||||||||:::|||||||||||||||||||
 999 LeuMetGluHisSerGlnLeuSerLeuSerGlyIleIleLeuIleLeuIleSe 1015
2982 CTGTGAAAATGACCTTCATTTATGCCGGAGAATAATTTTGCCAGGACATAG 3031
          |||||||||||||||||||||||||||||||||||||||||||||
1015 rCysGluAsnAspLeuHisLeuCysArgGluTyrProPheAlaArgGlyIle 1032
3032 ATGTTCACAAFGCACAGAGTTCGTTCTGCACTGGAGTGTGTCACTCAAACGCTG 3081
          ||||||||||||||||||||||||||||||||||||||||||||||
1032 spValHisSerAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeu 1048
3082 GACTATGAATATAATAAAGTTTAAC 3105
          |||||||||||||||||||||||
1049 AspThrTyrGluSerTyrLysPheAsn 1056
```

---

alignment_scores:
         Quality: 3841.00   Length: 745
         Ratio: 5.163   Gaps: 1
Percent Similarity: 99.866   Percent Identity: 99.463

alignment_block:
us-09-664-641-10 x O15404

Align seq 1/1   to: O15404   from: 1   to: 744

```
 874 GAAAATGAGGAACAAGAGATTCTCAGAAATGAGGGGTAGTACAGAGATGAGAAGTC 923
          |||||||||||||||||||||||||||||||||||||||||||||
```

seq_name: sp_vertebrate:Q90WJ3

seq_documentation_block:
ID   Q90WJ3   PRELIMINARY;   PRT;   1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell. Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE   1256 AA;   144187 MW;   E7795B12C1A42DE1 CRC64;

alignment_scores:
                Quality: 3425.50   Length: 1332
                Ratio: 4.112   Gaps: 20
Percent Similarity: 62.558   Percent Identity: 53.829

alignment_block:
US-09-664-641-10 x Q90WJ3

Align seg 1/1 to: Q90WJ3
165 ... from: 1 to: 1256

165 CCTTCAAAATCACCTCTAGGACACAGGTTATTCAGCTTCTCAAGGCTGGAAA 214
||| |:|:|| ::|||||||||||||||||||||||||||||||
20 ProValGlyAspLeuAspProLysValValGlnMetLeuMetLysAlaGlyGly 36

2723 AGCAGAACTACAATTCTCCGAGAATGCTGAGGCAGGAAGTACTTTTGTCTTTTC 2772
|||||||||||||||||||||||||||||||||||||||||||||||
617 luGlnAsnTyrIleLeuArgAspAlaGluAlaGlyValLeuPheSerPhe 633

2773 AGCTTGGGAAGAANTCCTTAAAAACGGGCACACGTTTCTCCACTCTTTAAGGC 2822
|||||||||||||||||||||||||||||||||||||||||||||
634 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 650

2823 AAAATATTTTTCATTACACCACCTGGAAATCTGCCCAAGTCTTTTCCACTATGA 2872
|||||||||||||||||||||||||||||||||||||||||||||
650 aLysTyrPheTyrIleThrProIleCysProSerLeuSerThrMetL 667

2873 AGGCAATCGTAGAGTGGTGCAGGAGGAGAAAGGTGTTATCCAAGCAGCCATCT 2922
|||||||||||||||||||||||||||||||||||||||||||||
667 ysAlaIleValGluCysAlaGlyGlyLysValLeuSerLysGlnProSer 683

2923 TTCCGGAAGCTCATGGAAGACACAGCAGACAAACTTGCAGTTGTCGGAAATAAT 2972
|||||||||||||||||||||||||||||||||||||||||||||
684 PheArgLysLeuMetGluHisLysLysGlnAsnSerSerLeuSerGluIleI 700

2973 TTTAAATATCCTGTGAAAAATGACCTTCAATTATGCCGAGAATATTTTGCCA 3022
|||||||||||||||||||||||||||||||||||||||||||||
700 eLeuIleSerCysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaA 717

3023 GAGGCAATAGAATGTTCACAAATGCACAGAGTTCGTTCTGACTGGGAGTGCTCACT 3072
|||||||||||||||||||||||||||||||||||||||||||||
717 rgGlyIleGluCysSerGlnMetHisArgValArgSerAspTrpGluCysSer 733

3073 CAAAACGCTGGACTATGAAATCANTATAAGTTTAAC 3105
|||||||||||||||||||||
734 GlnThrLeuAspTyrGluSerTyrLysPheAsn 744

215 ACGGAGGAGAAGTTTCCTACAATGCACTAGCCTTCACACAATAANTCTCAGAG 264
|||||||||||||||||||||||||||||||||||||||||||||
36 sAlaLysGluValSerIleArgAsnAlaLeuAlaThrHisIleIleLeuAlaGluA 53

265 ATGGGGGACAATCCCAGAGGTGGGAGGAGAAGTCGGGAAGAGTCTTTGACTTACCT 314
|||||||||||||||||||||||||||||||||||||||||||||
53 spGlyAspAsnProGluValGlyGlyGluAlaGluArgGluValPheAspLeuPro 69

315 GTTGTAAAGCCTTTCTTGGGTGGATTCTGTCCGTTCAGTGTGGAACTCTTCT 364
|||||||||||||||||||||||||||||||||||||||||||||
70 ValValLysProSerTrpValIleGlyPheCysProPheSerValValArgIle 86

365 GCCAGTAAATGGTTTTTCTCCAGAATCATGTCAGAATTTTTTGGAATCA 414
||| |||||||||||||||||||||||||||||||||||||||||
86 uProGluAsnGlyPheSerProLeuSerGlyInIlePhePheGlyValT 103

415 CTGCCTGCCTTTCTCAGGGTGTGATACAAGCTGGAGTCTTTGTTGGAG 464
|||||||||||||||||||||||||||||||||||||||
103 hrAlaCysGlyLeuSerGln.Val. 109

465 TCTTCCAGAGCTCTCCAAGGAGGTAGGGAAGGGAGCGTTGGTGCCAGCAG 514
|||||||||||||||||||||||||||||||||||||||||
109 109

515 AAGTTGGGAAGCAAGAAGAAGATCACATCTGCCTTCTTCTGTGACCGGGTTANTGATG 564
|||||||||||||||||||||||||||||||||||||||||
109 109

565 CAGGCTGAGGCCTCTGTCGTTGTTAAANTGCTGGGGTGTCGTCATCTGAAGAGCAGG 614
||| ||| |||||||||
115 ...SerProAspSpArgGas 115

615 TGCCCTGTGGGCGTTGGTTCATTCATGGGGAGATTGGCCAGCTPAACCC 664
|||||| ||| |:::||||||||||||||||||
132 nSerLeuTrpProAlaLeuThrThrPheTyrGlyIleGlyAspCysGlnLeuSerI 132

665 TCAATAAGAAAATGCACGCACATTGANTGTTCCAGAGCCACCAAAGGGGGGGAAA 714
|||| ||||||||||||||| :||:||:|||||||||||
148 euAsnLysLysCysThrHisLSerIleValProGluProLysGlyAsnLys 148

715 TACGAAATGTGCTTTAAAGCGGAGCGAAGTANTAAANTGACTCCCGACTG 764
|||||||| ||||:::::||||||||||||||||||
165 TyrGluTyrAlaPheGlnArgGlySerIleLysValThrProAspTr 165

765 GGTTTCCTGGATTGCGTATCAGAAAAACCAAAAAGGACGCAACCCATTTTANTC 814
||||:::||||||||||||||||||||||||||||||||||||||
182 isProArgLeuIleValIleTyr......GluGluGluGluGluGluGlu 196

815 ATCCTCCTCTGATTANTTATTTATGAAGAGGAAGAGGAGAGGAGAAGGAGGAG 864
|||| ||||||||||||||||||||||||||||||||||||||||
197 GluGluValAspSerValSerGlyLysThrLysLysSerAspGluALeuTyrH 213

915 TGAG......... AAGTCAAGCCCTCCCAGCCTTCAAGAAGGGTCTCCCTT 955
||:: |||||||||||||||||||||
213 pLysTyrSerSerIleSerSerProValThrSerArgAspProlySerProV 230

956 CAGGTGACCAGCAGTGTTTTCACCTAA...TCCAACACTGAAAAANTCTAA 1002
||| ||||||||||||||| :::||: ||| |||
230 alHisAsnArgGluSerProLysArgSerSerSerAspLysIleLys 246

1003 GGGGAATTAANTGTTTGANTGANTTCTCAGANTCATCACCGAAAAAACAGGA 1052
:::||||||||||||||||||||||||||||||||||||||||
247 SerGluLeuMetPheAspAspSerSerSerProGluGlyLysGly1 263

1053 GAGAAANTTTAAACTGGGACCCCGGGCGAAGTCCCACAGTAGCTCGACAA 1102
||||||||||||||||||||||||||||||||:||:||
263 uArgAsnLeuAsnThrPThrProAlaGlyValProGlnMetSerThrAlaL 280

1103 AACGCAGGCTGCCTCAGGGAGAAAGAGCCTGGGTTTGGTTGANTAACTGTGTGCC 1152

OM of: US-09-664-641-10 to: SPTREMBL_19:*　　out_format : pfs

Date: Jun 4, 2002 5:49 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p_model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04052002_110933_9250/app_query.fasta_1.3697
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext_HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER-US09664641_eCGN1_1312 -NCPU=6 -ICPU-3 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXY -WAIT -THREADS-1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 230.300000

Score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_rodent:Q9Z0W6	+	4059.50	5875.12	0.0	1056	Q9Z0W6 mus musculus (mouse). F
sp_human:O15404	+	3841.00	5561.10	0.0	744	O15404 homo sapiens (human). ca
sp_vertebrate:Q90WJ3	+	3425.50	4951.40	1.5e-268	1256	Q90WJ3 xenopus laevis (african
sp_human:Q96HP2	+	2039.00	2947.67	1.4e-156	391	Q96HP2 homo sapiens (human). un
sp_invertebrate:Q9VUB6	+	852.00	1205.56	4.6e-60	1798	Q9vub6 drosophila melanogast
sp_invertebrate:Q94046	+	524.00	734.07	1.4e-33	1074	Q94046 caenorhabditis elegan
sp_invertebrate:Q9Q370	+	524.00	734.03	1.4e-33	1076	Q9u370 caenorhabditis elegan
sp_human:Q1676	+	342.50	463.12	8.9e-19	2089	Q1676 homo sapiens (human). k
sp_human:Q96OC2	+	340.50	460.21	1.3e-18	2090	Q96oc2 homo sapiens (human). k
sp_plant:Q91VF7	+	316.50	432.68	8.9e-17	1041	Q91vf7 arabidopsis thaliana. un
sp_roden:Q91VX3	+	283.50	388.38	4.5e-14	1296	Q91vx3 mus musculus. s
sp_plant:Q9zR18	+	265.50	361.78	1.1e-12	765	Q9zr18 arabidopsis thaliana (mc
sp_human:Q9FB9	+	263.50	358.22	2.0e-12	1435	Q9eb9 homo sapiens (human). p
sp_human:Q92547	+	263.50	351.40	2.0e-12	1550	Q92547 homo sapiens (human). k
sp_fungi:Q9P5J5	+	229.00	307.74	8.0e-09	838	Q9pj5j5 neurospora crassa. cons
sp_invertebrate:Q9VFR7	+	225.50	303.46	1.9e-09	775	Q9vfr7 drosophila melanogast
sp_invertebrate:O24163	+	220.50	291.76	5.5e-09	1180	O24163 drosophila melanogast
sp_invertebrate:Q9VED3	+	220.50	291.71	5.5e-09	1186	Q9ved3 drosophila melanogast
sp_invertebrate:O44011	+	219.50	288.00	7.2e-09	1457	O44011 dictyostelium discoid
sp_invertebrate:Q9W4E8	+	216.00	293.94	9.6e-09	517	Q9w4e8 drosophila melanogaste
sp_plant:Q93VQ3	+	215.50	283.70	1.3e-08	1272	Q93vq3 oryza sativa (rice). hy
sp_invertebrate:O76853	+	214.00	293.23	1.3e-08	1420	O76853 dictyostelium discoide
sp_fungi:P79025	+	214.00	291.75	1.4e-08	483	P79025 emericella nidulans (asp
sp_virus:Q08148	+	214.00	282.47	1.9e-08	1162	Q08148 kaposi's sarcoma-associ
sp_vertebrate:Q90Y12	+	213.50	274.54	2.9e-08	1822	Q90yi2 fugu rubripes (japanese
sp_invertebrate:Q90VC9	+	212.00	283.02	2.4e-08	838	Q9vc9 dictyostelium discoid
sp_invertebrate:O18346	+	210.50	278.87	3.4e-08	1010	O18346 drosophila hydei (fru
sp_virus:Q9O71	+	210.00	276.96	3.9e-08	1129	Q9qr71 kaposi's sarcoma-associ
sp_invertebrate:Q9VMS2	+	208.50	266.64	6.7e-08	2439	Q9vms2 drosophila melanogast
sp_invertebrate:Q95GC5	+	207.00	262.29	9.6e-08	2994	Q95gc5 dictyostelium discoid
sp_invertebrate:Q9V7C5	-	206.00	273.98	6.9e-08	925	Q9v7c5 drosophila melanogaste
sp_invertebrate:Q9YM8	-	206.50	268.20	8.4e-08	1598	Q95ym8 apis mellifera (honey
sp_invertebrate:Q17641	-	205.50	288.88	4.8e-08	197	Q17641 caenorhabditis elegans
sp_invertebrate:Q9VN59	+	205.00	271.70	9.2e-08	934	Q9vn59 drosophila melanogaste
sp_invertebrate:Q94485	+	204.50	274.64	8.9e-08	765	Q94485 dictyostelium discoide
sp_invertebrate:Q9NBT2	-	203.00	255.01	2.1e-07	3439	Q9n8t2 trypanosoma brucei. F
sp_human:O15410	+	202.50	271.86	1.3e-07	652	O15410 homo sapiens (human). ca
sp_human:O75557	+	202.50	259.89	1.9e-07	2023	O75557 homo sapiens (human). c
sp_human:Q99ND7	+	202.50	259.89	1.9e-07	2023	Q9und7 homo sapiens (human). o

sp_virus:Q9DUN0	+	201.00	265.41	2.0e-07	976	Q9dun0 kaposi's sarcoma-asso
sp_rodent:Q64507	-	200.00	279.24	1.4e-07	230	Q64507 mus musculus (mouse).
sp_human:Q99B44	+	200.00	268.27	2.0e-07	649	Q9b644 homo sapiens (human).
sp_virus:O49047	+	200.00	262.80	2.5e-07	1089	O49047 kaposi's sarcoma-ass
sp_invertebrate:Q9US87	+	200.00	255.74	3.1e-07	2123	Q9us87 dictyostelium disc

seq_name: sp_rodent:Q9Z0W6

seq_documentation_block:
ID　　Q9Z0W6;　　PRELIMINARY;　　PRT;　　1056 AA.
AC　　Q9Z0W6;
DT　　01-MAY-1999 (TrEMBLrel. 10, Created)
DT　　01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT　　01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE　　PAX TRANSCRIPTION ACTIVATION DOMAIN INTERACTING PROTEIN PTIP.
GN　　PAXIP1.
OS　　Mus musculus (Mouse).
OC　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　　Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX　　NCBI_TaxID=10090;
RN　　[1]
RP　　SEQUENCE FROM N.A.
RC　　STRAIN=FVB;
RX　　MEDLINE=20356635; PubMed=10908331;
RA　　Lechner M.S., Levitan I., Dressler G.R.;
RT　　"PTIP, a novel BRCT domain-containing protein interacts with Pax2 and
RT　　is associated with active chromatin.";
RL　　Nucleic Acids Res. 28:2741-2751(2000).
DR　　EMBL; AF104261; AAD17927.1; -.
DR　　MGD; MGI:1990430; Paxip1.
DR　　InterPro; IPR001357; BRCT.
DR　　Pfam; PF00533; BRCT; 6.
DR　　SMART; SM00292; BRCT; 5.
DR　　PROSITE; PS50172; BRCT; 5.
SQ　　SEQUENCE 1056 AA; 119331 MW; 99CAF4EF4EE53ED6 CRC64;

alignment_scores:
Quality: 4059.50　　　　　　Length: 1142
Ratio: 4.572　　　　　　　　Gaps: 11
Percent Similarity: 77.758　　Percent Identity: 70.928

alignment_block:
US-09-664-641-10 x Q9Z0W6

Align seg 1/1 to: Q9Z0W6　　from: 1　to: 1056

```
17  ................................................  17

33  CCGGAAATTCCCGGGTTCGACGGATTTCGTCGAGTGCAAGAGCGGCCTGC  82
    |||||||||:::::::::::::::|:|:|||::::||:::
6   ProGluValProGluGlyLeuPheArgGluValLys...............  17

83  TGAGGGCGCCCCGCCGGGCCGGGTCTGGGTCTGGGCCGGGGGAAACTCCGGGGACGGCGC  132
    ...

17  ................................................  17

133 CGCGACGCAGGGCTACGGATCTCTCGGCCCCCTGCCCCTTCAAATCACCCTCTAG  182
    |||:|||:|:|:   |||||
18  ..TyrTyrAlaValGlyAspIleLeuAspPro.................  26

183 GCACAGGTTATTCAGCTTCTCAAGGCTGGAAAAGGAAAGGAAAGTTCCTA  232
    ||||||||||:||||||||||:||||||||||||||||||
27  ..GlnValIleGluIleLeuLeuLysAlaGlyLysAlaLysGluValSerTy  42

233 CAATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAATCCAGAGG  282
    |||||||||||||||||||||||||||||||||||||||||
43  rAsnAlaLeuAlaSerHisIleIleSerGluAspGlyAspAsnProGluV  59

283 TGGGAGAAGCTGCGGGAGAGTCTTTGACTTACCTGTGTTGTAAAGCCTTCTTGG  332
    ||||||||||||||||||||||||||||||||||||||||||||||
59  alGlyGluGluAlaArgGluValPheAspLeuProValValLysProSerTrp  75

333 GTGATTCTGTT GTTCAGTGTGTGGAACTCTTCTGCCCAGTAAATGGTTTTTC  382
    ||||||||||||||||::::||||||||
```

```
seq_name: sp_human:O15404

seq_documentation_block:
ID   O15404      PRELIMINARY;      PRT;   744 AA.
AC   O15404;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   CAGF28 (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
DR   EMBL; U80735; AAB91434.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 2.
FT   NON_TER      1      1
SQ   SEQUENCE   744 AA;  83747 MW;  8FB1D5E267A70E0 CRC64;

alignment_scores:
           Quality:  3841.00             Length:   745
           Ratio:     5.163              Gaps:       1
  Percent Similarity:  99.866   Percent Identity:  99.463

alignment_block:
US-09-664-641-10 x O15404

Align seg 1/1  to: O15404  from: 1  to: 744
```

seq_name: sp_vertebrate:Q90WJ3

seq_documentation block:

```
ID   Q90WJ3          PRELIMINARY;     PRT;   1256 AA.
AC   Q90WJ3;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SWIFT.
GN   K14.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=21260044; PubMed=11359898;
RA   Shimizu K., Bourillot P.Y., Nielsen S.J., Zorn A.M., Gurdon J.B.;
RT   "Swift is a Novel BRCT Domain Coactivator of Smad2 in Transforming
RT   Growth Factor beta Signaling.";
RL   Mol. Cell. Biol. 21:3901-3912(2001).
DR   EMBL; AF172855; AAK55123.1; -.
SQ   SEQUENCE   1256 AA;  144187 MW;  E7795B1C1A42DE1 CRC64;
```

alignment_scores:

```
       Quality:   3425.50        Length:        1332
         Ratio:      4.112        Gaps:            20
Percent Similarity:  62.538   Percent Identity:  53.829
```

alignment_block:

US-09-664-641-10 x Q90WJ3

Align seg 1/1 to: Q90WJ3   from: 1   to: 1256

```
 280  ysArgGlnLeuHisGlnGly..............................
      |||||::|||                        |||||
                                                       286

1153  AATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGGTAA 1202
      :::||:|||||
 287  .........AlaProGly.                                289

1203  TTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAAATGA 1252
      ||||||||  ::||||
 290  ..........SerGluArgProAspMetM                      296

1253  TAGCTACCTGGAGTTCAGCCTGTACCGGACAACTGAGGAATATTACTAAAT 1302
      :::|
 296  etAlaPheArgSerProAlaValArgThrLeuArgAsnIleThrAsnSer  312

1303  GCTGACATTCAGCAGATCGAACCGGCCCCATCAAAATGTACCACATATCTTA 1352
      |||||:::||||||||||||||||||||||||||||||||||||||:::
 313  AlaAspValGlnValAsnArgProSerAsnValAlaHisIleLeuGl     329

1353  GACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAAATCACAGCC 1402
      |||||||||||:::||||||||||||||||||||||||||||||||||||
 329  nThrLeuSerAlaSerThrLysSerLeuGluGlnArgValAsnHisProG  346

1403  AGGAGGGCAACTACAAAATGCCAATGGATGCCTGTCTGTTTACCCAAGTGAAA 1449
      ||||||||||
 346  nGlnGlyHis...   ProAsnAlaValLeuPheGlyGlnValLysPro   360

1450  GTCACTCCACAGACA..CACATGCTACAGACAG.........CACCA      1484
      :::|| |||:|||
 361  LeuThrSerGluAlaGlnHisLeuIleGlnSerHisGlnProHisHi     377

1485  GCAGGGCCCAGCACGCAGCAGCAGCAGCAGCAGCACCCGGTTTTACCTTCAGCCCC 1534
      :::||  ||
 377  sGlnGlyGlnGlnGlnGlnGlnAsnHisProLeuValGlnLeuGlnProG  394

1535  AGCAGGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAG            1569
                             ||||:::||||||
 394  lnGlnLeuLeuGlnLeuGlnGlnGlnGlnHisPheGlnGlnHisAlaLeuGlnGln 410

1570                              CAGATCTCAGCAA.........   1584
                                  |||:||||||||
 411  PheProGlnHisGlnPheProGlnValAsnGlnGlnHisProPheThrGl  427

1585  ...CCTTACCCCAGCAGCCGCGCATCCA........               1611
      :::|| ||||||||||||
 427  nLeuGlnPheProGlnlnGlnLeuGlnIleuHisArgp             444

1612  .............TTTTCACAGCAG.......                   1623
                   ||:::|||||
 444  roGlnGlnThrIleGlnHisPheGlnGlnHisAlaLeuGlnGln       460

1623  .........................                          1623
 461  GlnLeuHisGlnLeuGlnGlnLeuGlnHisLeuGlnProLysProGlnThrLe  477

1623  .........................                          1623
 477  uGlnGlnAsnMetGlnGlnAsnLeuGlnProAsnLeuGlnI          494

1623  .........................                          1623
 494  leLeuGlnHisGlnGlnThrIleGlnThrProSerGlnGlnGlnAla     510

1623  .........................                          1623
 511  LeuGlnProAlaIleGlnGlnMetLeuGlnProAsnIleGlnGl       527

1623  .........................                          1623
 527  nGlnGlnThrLeuGlnSerAsnLeuGlnltnGlnThrLeuGlnProIleI  544
```

```
 544  leGlnGlnGlnGlyLeuGlnGlnGlnAsnLeuGlnSerLeuGln        560

1623  ...........................                        1623
 561  GlnIleGlnGlnGlnMetGlnHisLeuPheThrProGlnIleysGlnGlnIl  577

1624  .CAGCAGCAGCAAACAGCAG........                       1641
      |||||||||||||||
 577  eGlnGlnGlnGlnGlnGlnMetLeuGlnLeuGlnSerGlnS          594

1641  ...........................                        1641
 594  erLeuGlnGlnGlnGlnMetGlnThrHisValLeuGlnGlnGlnGlnIle  610

1641  ...........................                        1641
 611  GlnThrGlnHisAlaLeuGlnGlnIleGlnAsnGlnAlaLeuGlnGlu    627

1641  ...........................                        1641
 627  nGlnProGlnGlnAsnGlnValGlnThrHisAlaLeuGlnHisArgl    644

1641  ...........................                        1641
 644  euGlnSerGlnThrLeuGlnGlnHisHisValGlnGlyGlnThrHrVal   660

1641  ...........................                        1641
 661  GlnGlnGlnThrHisGlnLeuGlnThrGlnGlnLeuGlnGlnHisGl     677

1642  ...CCACCACCATCG..........                          1653
           :::|||
 677  nIleGlnThrLeuGlnIleProHisGlnIleProAlaProAsnGlnHn   694

1653  ...........................                        1653
 694  isGlnIleProGlnMetLeuGlnGlnThrLeuGlnGlnGln          710

1653  ...........................                        1653
 711  GlnMetGlnProGlnIleGlnMetGlnSerGlyValGlnGl          727

1653  ...........................                        1653
 727  nGlnSerSerLeuGlnProGlnIleMetGlnGlnHisAsnLeu         744

1653  ...........................                        1653
 744  lnGlnValGlnHisGlnLeuGlnLeuGlnGlnArgMetGlnGln        760

1654  ...............CCTCAGCA                            1661
                     ||||||
 761  ArgGlnMetAlaLeuProGlnGlnIleAlaAsnGlnGlnProGlnPr    777

1662  GCATCAGCTTTTGGACANTGAICCAGCACAGTGGAGATTCCAGAGAAGGCT 1711
      |||||||     ||||::::|||||||
 777  oHisGlnLeuHisGlyHisAspProSerValGlnIleProGluProAsp  794

1712  TCTTATTGGGATGTGTTGTTTGCAAITGCGATTATCCGAGAGCAGATGTCT 1761
      :::|||||||||||:::|||||||||||||
 794  heLeuLeuGlyCysValPheAlaIleAlaAspTyrProGluGlnMetPro 810

1762  GATAAGCCAACTGCTGGCCACCTGGAAAAAGGAATAAATCCAGGACACATGGCG 1811
      ||||||||||||||||||||
 811  AspLysGlnLeuLeuAlaThrTrpLysArgIleLeuGlnProHisGlyGl 827

1812  CACTGTTGACCCACCTTCACGAGTCAGATGACACGCACACCTTCTGTGTAGGA 1861
      :::||::::|||||||
 827  yThrValAspIleAlaIleLeuSerSerArgCysThrHisLeuCysGlyl 844

1862  GTCAAGTCACGACGGCGTTATCCACACAGGCAAATAAGACAGAAAAGAGATGT 1911
      |||||||||||||||||||||||
 844  erGlnValSerSerMetTyrAlaGlnAlaAlaLeuLeuLysGluArgLysArgCys 860
```

```
1912 GTTACTGCACAGTCGGTTAAACACAGTCTTAAAAAAAAAAATGTACC 1961
     |||||||||||||||||||||||||||||||||||||||||||||||
 861 IleThrAlaHisTrpLeuAsnAlaValLeuLysLysLysMetValPr 877

1962 CCCCACCACCAGCCCTTCACTTCCCAGTGGCCTTCCCACCACGAGGAAAGC 2011
     ||||:::|||||||||||||||||||||||||||||||::::::||||||
 877 oProTyrArgAlaLeuHisPheProValAlaPheProGlyGlyLysP 894

2012 CATGTTCACAGCACATATTATTTCTGACTGTGATTTGTTGTTGATAGTGACAGA 2061
     ||||||||||||||||||||||||||||||||||||||||||||||||
 894 oProTyrArgAlaLeuHisPheProValAlaPheProGlyGlyLysP 894

 877 oProTyrArgAlaLeuHisPheProValAlaPheProGlyGlyLysP 894
```

alignment_scores:
```
           Quality:   2039.00          Length:     391
             Ratio:      5.215            Gaps:       0
Percent Similarity:  100.000   Percent Identity:  99.744
```

alignment_block:
US-09-664-641-10   to: Q96HP2

Align seg 1/1   from: 1   to: 391

```
2133  CCTCATCTGTAAAGAACCAACTGGTTTAAAGTTGAAAAAGCCAAAGAGT  2182
      ||||||||||||||||||||||||| ::|||||||:|||||||||||||
67    IleuIleCysIleLysGluProThrGlyLeuLysThrVrGlyLysAlaLysGluT  84

2183  GGAGGATACCCTGTCTCAACGCCCAGTGCCTGGCCGACATTCTTGCGGA  2232
      ||||||||||||||||||||||||||||||||||||||||||||||||
84    rpArgIleProCysValAlaSnAlaGlnTrpLeuGlyAspIleLeuGly  100

2233  ATGCTTGGAGAGTTCCTTAAAGTGTCTGCAGAGTTCTTGATGAGTATA  2282
      ||||||||||||||||||||||||||||||||||||||||||||||
100   spAlaTrpArgValProLeuLysValSerAlaGluLeuLeuMetSerIle  117

2283  TCTCCAGATCCATTTGCCCCTACCCAGCATTTAGTTTAAATCTTTTAG  2332
      ||||||||||||||||||||||||||||||||||||||||||||||
117   rLeuGlnAspProPheAlaProThrGlnHisLeuValLeuAsnLeuLeuA  134

2333  ATGCTTGGAGAGTTCCTTAAAGTGTCTGCAGAGTTCTTGATGAGTATA  2382
      ||||||||||||||||||||||||||||||||||||||||||||||
134   rLeuGlnAspProPheAlaProThrGlnHisLeuValLeuAsnLeuLeuA  150

2383  AGACTACCTTCCCAAACTGAAACAGAATGAAGTAGCTAGTTGTCCAGCCTTC  2432
      ||||||||||||||||||||||||||||||||||||||||||||||
150   ArgLeuProProLysLysGlnAsnGluValAlaAsnValGlnProSe  167

2433  TTTCCCAAAGACCCAGAATTGAAGAGTCACCCTCCCACTAAAAAGCTAA  2482
      ||||::|||||||||||||||||||||||||||||||||||||||||
167   rSerLysArgAlaArgIleGluAspValProProProThrLysLysLeuT  184

2483  CTCCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTGTCCAG  2532
      ||||||||||||||||||||||||||||||||||||||||||||||
184   hrProGluLeuThrProPheValLeuPheThrGlyPheGluProValGln  200

2533  GTTCAACAGTATATTATTAAGAAGCTCTACATTCTTGGTGGACAGGTTGCGGA  2582
      ||||||||||||||||||||||||||||||||||||||||||||||
201   ValGlnGlnTyrIleLeuLysLysLeuTyrIleLeuGlyGlyGlnValAlaGl  217

2583  GTCTGCACGAAGTGCACACCACTCATTGCCAAAAGTAGACTCGCACCG  2632
      ||||||||||||||||||||||||||||||||||||||||||||||
217   uSerAlaGlnIlyrCysThrHisLeuIleAlaSerLysValThrArgThrV  234

2633  TGAAGTTCCTTCGACGGCCAGTTTCTCTGCTGAAGCACATAGTCGACGCCAGAG  2682
      ||||||||||||||||||||||||||||||||||||||||||||||
234   aLLysPheLeuThrAlaSerVaLValLysHisIleVaLThrProGlu  250

2683  TCGCTCGGAAAGAATGCTTCAGGTGTCTCAGAAGTTCATTCATGAGCCAGCAGACTA  2732
      ||||||||||||||||||||||||||||||||||||||||||||||
251   TrpLeuGlyLvCysPheArgCysGlnLysPheLIeAspGluGlnAsnArg  267

2733  CAATCCTCCGAGATCCTGAGGCAGAAGTACTGTTTCTCTTTCAGCCTTGGAAG  2782
      ||||||||||||||||||||||||||||||||||||||||||||||
267   rIleLeuGluArgAspAlaGluVaLLeuLeuPheSerPheSerLeuGluG  284

2783  AATCCTTAAAAACGGGCACACCGTTCTCCACTCTTAAGCCACAAAATATTTT  2832
      ||||||||||||||||||||||||||||||||||||||||||||||
284   rIleIleLysSerGlyHisSerValSerProLeuPheLysAlaGlyTyrPhe  300

2833  TACATCACACCTGGAATCTGCCCAAGTCTTTTCCCACTATCAAAGCCAAATCGT  2882
      ||||||||||||||||||||||||||||||||||||||||||||||
301   TyrIleThrProGlyIleCysProSerLeuSerThrMetLysLysAlaLeu  317

2883  AGAGTGTGCAGGAGGAAAGGTGTTATCCAAGAGCAGCCCATCTTTCCGGAGC  2932
      ||||||||||||||||||||||||||||||||||||||||||||||
317   lGlyuCysAlaGlyGlyLysVaILeuSerLysSerSerProSerAlaAla  334

2933  TCATGGAGCACACAGCAGAACTCGACTTTCTCCGGAATAATTTTAATATCC  2982
      ||||||||||||||||||||||||||||||||||||||||||||||
334   euMetGluHisHisSerGlnAsnSerThrPheSerGlyAsnAsnPheAsnI  350

2983  TGTGAAAATGACCTTCATTTATGCCGAGAATATTTTTGCCAGAGGCAATAGA  3032
      ||||||||||||||||||||||||||||||||||||||||||||||
351   CysGluAsnAspLeuHisLeuCysArgGluTyrPheAlaArgGlyIleAs  367

3033  TGTTCACATGCAATGCAGAGTTCGTTCTGACTGGAGTGTGCTCACTCAAACGCTGG  3082
      ||||||||||||||||||||||||||||||||||||||||||||||
```

seq_name: sp_invertebrate:Q9VUB6

```
3083  ACTATGAATCATATAAGTTTTAAC  3105
      ||||||||||||||||||||||||
367   pValHisAsnAlaGluPheValLeuThrGlyValLeuThrGlnThrLeuA  384

384   spTyrGluSerTyrLysPheAsn  391
```

```
alignment_scores:
            Quality:  852.00              Length:     848
              Ratio:    1.906             Gaps:        24
   Percent Similarity:  52.712    Percent Identity:  28.774

alignment_block:
   US-09-664-641-10 x Q9VUB6

Align seq 1/1  to: Q9VUB6   from: 1  to: 1798  ..
```

```
 925 AGCGCTGCCAGCTTCTCAAGAAACGGTCTCTCCTTCAGGTGAGCCACCAGTTTTC  974
     ::|||:::||:|:: ::|||||    |||| ||:::|  ||||| |||| ||
 974 SerProIleGlnSerGlnProGlyAlaGlyProGlnGlnGlnSe  990

 975 ACCTAAAATTCCAACACTGAAAAATTCTAAAACGGGCAATTAATGTTTCATGATT 1024
     |||:||::|||::||||||::|||||||||::||||| |||||||||||
 990 rProGlnGlnGlnProPheAsnProAsn............AspProA 1002

 990 rProGlnGlnGlnProPheAsnProAsn............AspProA 1002

1075 GCCCAA.........CTCCCACAGTTAGCTGCACACAAAACGCAGGCGTGCC 1115
     :::                |||||||||||||||||||||||
1018 snGlnGlnMetLeuMetGlnArgGlnGlnGlnProGlyAlaGlyProGlnGlnIlePro 1018

1002 snGlnGlnMetLeuMetGlnArgGlnGlnGlnProGlyAlaGlyProGlnGlnIlePro 1018

1025 CTTCAGATTCATTCACCGGAAAAAACAGGAGAGAAATTTAAACTGGACCCG 1074
     :::||||:||||||||||:::|||   :||||::||| :::|:|
1018 snGlnGlnMetLeuMetGlnArgGlnGlnGlnProGlyAlaGlyProGlnGlnIlePro 1018

1116 TCAGGGAAAAGCAGCCTGGGTTGCATT............ 1140
     |||                     ||||:|||||||
1035 eThrGlyProAsnGlyGlnLeuIleGluGlnLnThrIleValGlnGlnP 1052

1035 eThrGlyProAsnGlyGlnLeuIleGluGlnLnThrIleValGlnGlnP 1052

1141 ...........AACTTGTGTGGCCAATGTCCCACCGGTCCCA...........GGTAAC 1176
        ::|||  ::|||||| |||||||||||||||| :||::|
1052 roGlnGlnGlnSerAlaProGlyThrProThrProThrProGlyAlaGlyGly 1068

1052 roGlnGlnGlnSerAlaProGlyThrProThrProThrProGlyAlaGlyGly 1068

1177 ATTTTGCCCCTTGCAGTTCGGCGGCTTAATTTAATGCCTGCTGGACAAAACCT 1226
        ||||||:||||  |||||| |||||||||||||:|||||||||
1068 GlyAsnAlaAsnIleMetThrGlnThrIle 1081

1069 VallueuPro.........GlyAsnAlaAsnIleMetThrGlnThrIle 1081

1227 C...CAAAGTTCTCGAAAGATCAGAAATGACAAATGCATAGCTACCTGGACTCCAGCTG 1273
     :::   |||||||||||||||||||||||||:||||||||||||
1081 uValMetThrSerThrThrProAspGlyGlnProGlnGlnInSerProGlnG 1098

1081 uValMetThrSerThrThrProAspGlyGlnProGlnGlnInSerProGlnG 1098

1274 TACGGACACTGAGGGAATAATTACTAATAATGCTGACATTCAGACGATGAAC 1323
     :::||| || :::::|||||| |::|||| | |||||||||||||
1098 lnMetAsnLeuLysThrLysThrAlaLeuAlaAsnMetLeuSerAsnArg 1114

1098 lnMetAsnLeuLysThrLysThrAlaLeuAlaAsnMetLeuSerAsnArg 1114

1324 CCGCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
     |||||  :||| |||||| |||||||:|||:|:|||||||||||||||
1114 GlyAsnAsnGlyAlaGlnThrProGln 1124

1115 LeuGlyAsnAsnGlyAlaGlnThrProGln 1124

1374 AAAATTTAGAACAGCAGGGTGAATCACAGCCACCAGGACACTACAAAATGCCA 1423
     :::|||:|||:|||||: :||||||  ||||||||||||||
1124 ...............GlnGlyGlnLeuLeu......pGlyGlnGlnProGlnG 1138

1125 ...............GlnGlyGlnLeuLeu......pGlyGlnGlnProGlnG 1138

1424 ATGGGAGTTGGTCGTTTAGCCAAGTGAAAGT......CCA 1458
     :::||| ||||||| | ||||||  ::|||| :::||
1138 lnGlnGlnIleValGlnIleValValThr.GlyAlaIleGlnIleGln 1154

1138 lnGlnGlnIleValGlnIleValValThr.GlyAlaIleGlnIleGln 1154

1459 GAGACACACAATGCTTACAGCAGCAGCACCAGCAGCAGCAGCA 1508
     |||||||||||| |||||||||:|||||||||||||||||||
1155 GlnGlnGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnCl 1171

1155 GlnGlnGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnCl 1171

1509 G............CACCCGGTTTTACACCTTCAGCCC......... 1533
     |                 ||| |||:||:|||||
1171 nMetValGlnGlnGlnLeuGlnProProGlnGlnHisValGlnGlnP 1188

1171 nMetValGlnGlnGlnLeuGlnProProGlnGlnHisValGlnGlnP 1188

1533 ...............CACCAGATAATGCAGCTCCAGCA 1556
        ::::||:|::|||||||||||||
1188 erAlaAlaGlyAlaLeuArgMetMetGlyGlnGlnHisAsnAlaThrAla 1204

1188 erAlaAlaGlyAlaLeuArgMetMetGlyGlnGlnHisAsnAlaThrAla 1204

1205 GlyValProGlyProProArgThrGlnGlnGluLeuLeuMetLeuGlnGl 1221
     |||:|||| |||||||||::|||||||||:|||||||||||||||
1534 ..........AGTGCTACACGGCATTCAGTCGCCAGGATCCATTG 1556
```

```
1557 GCAGCAGCAGCACGAGATCTCTCAGCAA.................... 1584
     |||||||||||||||:::    ||||||
1238 laLeuProGlnGlnGlnGlnAlaProGlnGlnHisLeuValGlnGlnGln 1254

1221 nClnGlnGlnLnLeuLeuGlnGlnLnMetArgArgValAlA 1238

1585 ..CCTTACCCCAGCAGCAGCCCGCCGCCATCCATTT............ 1614
        |||||||||||||||||  |||||
1255 IleLeuValAlaAlaProThrGlnGlnSerProGlnGlnGlnMetAlaLeuAlaVa 1271

1238 laLeuProGlnGlnGlnGlnAlaProGlnGlnHisLeuValGlnGlnGln 1254

1615 .............TCACACGACAGCAGCAGCAGCAA.......... 1635
                        ||||||||||||||||
1271 lGlyAlaGlyValGlyValAlaGlnProValGlnArgGlyThrProHisGlyThrTyrIleG 1288

1635 ................. 1635

1271 lGlyAlaGlyValGlyValAlaGlnProValGlnArgGlyThrProHisGlyThrTyrIleG 1288

1636 ..............CAGCAGCCACCACCATGCCTCAGCAGCATCAGCTT 1671
                     :|||||||||||||||||||||||
1288 InThrArgProGlyGlnGlnProIleProMetPro...........GlnPhe 1301

1288 InThrArgProGlyGlnGlnProIleProMetPro...........GlnPhe 1301

1672 TTTGGACATGATCCACGCAGTGGGAGATTCCAGAAGAAGCCTTCTTTATTGGG 1721
     :::||||:||||| :|||||| ||||||||:::|||  ||||||:||||
1302 TyrGlyHisAspProAsnLeuGlySerProLeuAlaAsnProGluPheLeuValGl 1318

1302 TyrGlyHisAspProAsnLeuGlySerProLeuAlaAsnProGluPheLeuValGl 1318

1722 ATTGTGTTTGCAATTGCGGATTATTCCAGAGCAGAGATGTCTGATAAGGAAC 1771
     ||||:|||::|||:|| ||||||||:||||||||||||:||||::|||
1318 yCysThrPheTyrIleValGlnGluTyrAspGluThrAspGlyAlaSpGlu.... 1333

1318 yCysThrPheTyrIleValGlnGluTyrAspGluThrAspGlyAlaSpGlu.... 1333

1772 TGCTGGCCACCTGGAAAGGATAATCCAGGCACATGGCGGCACTGTTGAC 1821
     |:||||:|||| :|||||||||||||||||||||||||||||||||
1334 .LeuProIleTrpLeuAspThrIleArgGlnPheGlyGlyAspIleGlu 1349

1334 .LeuProIleTrpLeuAspThrIleArgGlnPheGlyGlyAspIleGlu 1349

1822 CCCACCTTCACGAGTCCCAGCATGGAGTTCCACCTTCTTCTGTGGAGTCAGAAGTCAG 1871
     :::     ||||||||| ||||||||||:|| |||||||||||||:||
1350 ArgValTyrCysProArgValThrHisValIleIleCysSerArgThrGlnArgHi 1366

1350 ArgValTyrCysProArgValThrHisValIleIleCysSerArgThrGlnArgHi 1366

1872 CAGCGCGTATGCACAGGCAATAAGAGAAAGGAAAGAAAGAGATGTTGATTACTCCAC 1921
     |||||||||| :|||| |||||||||||||||||||||||||:|||::|
1366 sGlyValValMetGlnAlaLeuArgAspAlaLysArgArgCysValThrAlaT 1383

1366 sGlyValValMetGlnAlaLeuArgAspAlaLysArgArgCysValThrAlaT 1383

1922 ACTGGTTAAACACAGTCTTAAAAAAAAAAAAAGGTACCGCGACACCGA 1971
     :::||||||| :|| :::     ||||||||||||||||||:|||
1383 yrThrpLeuSerAspIleCysSerLeuLysSerGlnLeuMetProProGln 1399

1383 yrThrpLeuSerAspIleCysSerLeuLysSerGlnLeuMetProProGln 1399

1972 GCCCTTCACTTCCCCACTGGGCCTTCCCCACCAGGA.......... 2012
     :::  ||||||||||||||   ||||||||||||:GGAAAGCC
1400 ProLeuHisLeuPro........PheProSerGlnProHisGlyTyrTyrArgLysPr 1414

1400 ProLeuHisLeuPro........PheProSerGlnProHisGlyTyrTyrArgLysPr 1414

2013 ATGTTCACAGCATAATAATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAG 2062
     :||||||||||:|||||||  :||||||| |||||:|||||||||||
1414 oLeuGluArgTyrIleIleProThrSerGluGlyPheGluGlyGluValV 1431

1414 oLeuGluArgTyrIleIleProThrSerGluGlyPheGluGlyGluValV 1431

2063 ATGACCTAAAATTAATGCTTAATTGGCAGGTTGCCAAAATNAGGGGTTAT 2112
     :::   |||||||:::||| :|::||||| | ||||||||||:|||
1431 alArgLeuGlnLnMetAlaGluGlucysGlyAlaIleTyrThrSerTyr 1447

1431 alArgLeuGlnLnMetAlaGluGlucysGlyAlaIleTyrThrSerTyr 1447

2113 CTATGCCCCAGCAACAGTCCTCATCTTGTAAAGAACCAAACTGGTTTAAA 2162
     |||   :: |||||||||||||||:::||:||:|| ::||
1448 LeuSerLysValAsnThrValValCysLysGlnLeuGlyAsnLy 1464

1448 LeuSerLysValAsnThrValValCysLysGlnLeuGlyAsnLy 1464

2163 GTATGAAAAAGCCAAAGAGTGGAGGAGATACCCTGTGTCAACGCCCAGTGGC 2212
     |:::||||||||||||||||||||||:||| || || ||||||||||
1464 sPheAsnAlaAlaLysGlnTrpAsnIleProMetValAsnAlaLeuTrpL 1481

1464 sPheAsnAlaAlaLysGlnTrpAsnIleProMetValAsnAlaLeuTrpL 1481

2213 TTGGGGACANTCTTCTGGGAAACTTTGAGGCACTGAGGCAGATTCAGTAT 2262
     |:|:|| :::|||||| |||||||:::||| | |||||||:|||||||
1481 euSerAspValCysIleGlyAsnLeuSerGlyLeuSerGlnTyrGluAsn 1497

1481 euSerAspValCysIleGlyAsnLeuSerGlyLeuSerGlnTyrGluAsn 1497

2263 AGTGCTACACGGCATTCAGTCGCCAGGATCCATTGCCCCTACCACCAGCA 2312
     :::||| ||||| || |||:|||||||||||||||||||||||||
1498 ProLysTyrGlnGlnTyrAsnLeuValAlaProPheArgIleGluCysAs 1514

1498 ProLysTyrGlnGlnTyrAsnLeuValAlaProPheArgIleGluCysAs 1514
```

seq_name: sp_invertebrate:Q94046

seq_documentation_block:
ID   Q94046
AC   Q94046;

PRELIMINARY;   PRT;   1074 AA.

DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T13F2.3A PROTEIN.
GN   T13F2.3A.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RL   Submitted (OCT-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81122; CAB03354.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 4.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 4.
SQ   SEQUENCE 1074 AA; F52E1423226I0C50 CRC64;

alignment_scores:
             Quality: 524.00
             Ratio: 1.444
    Percent Similarity: 52.457   Percent Identity: 25.000

alignment_block:
US-09-664-641-10 x Q94046

Align seg 1/1   to: Q94046   from: 1   to: 1074

              Length: 692
              Caps: 24

```
 1647 ACCAtCGCCTCAGCAGCATCAGCTTTTTGGACAtGAtCCA............ 1686
      ||:|||||
  492 nThrMetAsnGlnMetLeuAsnGlnMetHis...HisAspValAsnProPheP 508

 1687 .....GCAGTCGAGATTCCAGACAGAAGGCTTCTTATTGGGATGTGTGTTT 1731
      :||||| ||| |||
  508 rolleGlnMetAsnValAspProGluTyrLeuThrGlyIleHisPhe 524

 1732 GCAATGCGGATTATCCAGAGCAGATGTCTGATAAGCAACTGCTGGCCAC 1781
      ||| ||||||
  525 HisCysPheAspSerGlulysysLeuPheGlulysLeuAspArgLysAs 541

 1782 CTtGAAAAGGAAtAATCCAGCCACAATGGCCGGCCACtGTTGAC........ 1822
      :::::: :: C
  541 nLeuGluPheMetIleLeuTyrHisGlyGlyAspIleGluPheGluAlas 558

 1823 CCACCTTCACGAGTCGA.......... TGCACGGCCACCTTCTCTGCTCTGAGAGT 1863
      |||
  558 erlysPheHisGluArgIleHisGlnIleThrHisValValAspSer 574

 1864 CAAGTCACGAGGCCGTGTAGCAATAAGAGGAGAAGAGAGAtGTGT 1913
      ||||||||||| ||
  575 CysArgAsnThrAsnValMetThrSerIleuGluHisArgLysArgIleVa 591

 1914 TACTCCACACTGCGTTAAAACACTCTTAAAAAAAAAATGGTACCGC 1963
      :::::: :
  591 lSerIleGlnTrpIleIleAspValMetGluArgGluArgValGluLeuP 608

 1964 CGCACCGAGCCCTTCACTCCCAGTGCGTTCCACCAGGAGGAAAGCCA 2013
      |||
  608 roTrpArgIleAlaHisIleuProProProAsnGluAsnPheArgPro 624

 2014 TGTTCACGAGCAtAATATTTCTGTGACTGGGATTCGTTAGTGACAGA 2063
      :::::::|| |||
  625 TyrleuGlylysLeuPheSerLeuSerGlyPheAspSerSerGluArgAl 641

 2064 TGACCTAAAATTAAtTGGCTTATTGGCCAGGTTGCCAAAATAtACGGGTTATC 2113
      : ||| |||
  641 aAlaIleSerPheMetALaGluMetMetGlyValAlaLysIleThrProPheL 658

 2114 TATCCCGACGAACACAGTCCTCATCtGTAAAGAACCAACCAACTGGTTAAAG 2163
      :::|||||| :|||
  658 cuALaArgGlnAsnAspLeuGluLeuAlalysGlnProAlaAsp... 673

 2164 TATGAAAAAGCCAAAGAGTGGGAGGAtACCCtGTGTCAACGCCCAGTGGCT 2213
      |||||:||||||
  674 ValThrArgArgALaGlnGluTrpGlulysValProValAlaAsnPheGlnTrpIl 690

 2214 TGGCCACAtTCTTCtCGGGAAACtTTGAGGCGCACTGCAGCCAG............ 2253
      :::|||||||| |||
  690 eALaAspALaIyrValCysAsp...GluAlaSerProGlnLysArgProA 706

 2254 .AtTCACTATAGTCGCCTACACCGCATTCAGTCCTCCAGGAtCCCATTGCC 2301
      :::|||| ||||
  706 snValGluLeuGluAsnProArgTyr......... GlnLeuGluGluAsnProValMet 719

 2302 CCTACCCAGCATTTAGTTTTAAAT............ 2325
      |||
  720 ThrAsnValAsnThrSerProAsnHisIleGluGlyAlaSerGlnIleGluPh 736

 2326 .......... CTTTTAGAtGCTTGGAGAGTtCCCTTAAAGtGTCTGCAGACt 2368
      : ||
  736 eGlyAlaMetMetSerCysTrpLysSerGlnThrIleLeuAlaAspGlnA 753

 2369 TGTTGAtGAGtATATAAAGACTACCTCCCAAACTGAAACAGAATGAAGTA... 2415
      :::||| |||
  753 laTyrGluGlyAlaArgGlnAsnArgLysThrLeuGlnAsnAsnThrPhe 769

 2416 .......... GCTAAtGTCCAGCCCTTCt........... 2433
      ||| ||:::
  770 PhePheProSerlysArgIleuSerAspGlnAlaValAlaProThrGluAs 786
```

```
 2434 .......... TCCCAAACAGCCAGAATTGAAACGTA.......... 2460
      ::: ||:||||||||||||::::::
  786 pGlulleLysSerSerArgAlaArgLeuAspGlulleLysThrHisTrpG 803

 2461 .......... CCACCTCCACTAAAAAGCTA... 2481
      ||| :::|||||||||||:::
  803 luGluHisGlnGlulysMetLysAspThrProLeuSerlysGluProTyr 819

 2482 .......... ACTCCAGAAtTGACCCCTTTGTGCtTTTC...ACTGGATt 2519
      ||| |||
  820 ArgValAlArgGluProGlnAlaArgValSerAlaIaTrpPheAspIyGluGlyIleIl 836

 2520 CGAGCCTGTCCAGGTTGCAAACAGTATTATTAAGAAGGCtTCTACAtTCTtGGTG 2569
      ::: ||||||||||:::::||||| :::|||||||
  836 eAspAspGluAlaLeuLeuLysLysLeuGluPheLeuGlyLysAsnIl 853

 2570 GAGAGGTTGCGGAGGTCGCAGAGAAGTGCCACACCTCAtTGCCAGCAAA 2619
      |||:|||||||| ::: ||||||||||:::|| :::|
  853 lyGluCysValGluLysIleArgAspAlaThrHisValValIleLeuSer 869

 2620 GTGACTCCGACACCGTGAAGTtCCTGACGGCGCGATTTCTtGTCGTGAAGCACAt 2669
      || ||||||||:::: |||
  870 GlyAlaArgSerLeuValLeuLeuGluSerIleIleLeuArgGlyLysAsnIl 886

 2670 AGTGACGCCAGAGTGGCTGGAAGAAATGCTTCAGGTCGtCAGAAGTtCAtTG 2719
      |:: ||||||||||||::: :
  886 eMetAspProGluTrpIleValAspSerTyrLysGlnLysMetTrpLeuA 903

 2720 ATGAGGCAGAACTACATCTCCGAGAtGCTGAGGCGAGAAGTTACtTTTCTCT 2769
      ||| :::|||:||||||::||||
  903 spThrGluAspTyrPheLeuHisLeuHisAspLysGluLeuGlulysGlyPheAla 919

 2770 TTCAGCTTGGAAGAtCCTTAAACAGGGCACACGTTTCTCACtCtCTTTAA 2819
      ::: |||||::|||| |||:::|||
  920 TyrAsnCys: ysArgSerValLeuValArgLeuArgALaArgAsnLysProValPheGl 936

 2820 GGCAAAAtAtTTTACAtCACACCtGGAAtCtCGCCCCAAGTtCTTTCCACtA 2869
      :||||||||| :::||| :::||||||||||||:::
  936 uAspIleGluPheHisValThrArgPheValGluProAsnGlnAsnAspL 953

 2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTAtCCAAGCAGCCA 2919
      :: |||||| :::|||||||||||::||||
  953 euValArgValLeuIleGluLeuGlyGlyAsnValHisSerGlulysPro 969

 2920 TCT.......... TTCCGGAGCtCAtTGGAGCACAGCACAGAAtCtCGAGtTT 2960
      ::: |||||||||:::|||
  970 AspProLysTyrLeuAlalysCysValGluLeuGluThrGluGln... 982

 2961 GTCGGAAAttAAtTTAAtTAtCCCtGTGAAAAtGACCTTCAtTTAtGCCGAG 3010
      |||||||:::||||||:::||||||||:::
  983 .......... ProPheIleIleIleSerCysSerGluAsnAspAlaArgPheLeuSerT 998

 3011 AAtATtTTCCAGAGCCAtAGAtGTTtCACAAtGCAGAGAGTTTCCTGACt 3060
      |||:||||||||| ::||||||:::||||||||||
  998 yrIleuAlaGlySerLyslleuProIleTyrAsnValAspLeuValLeuPhe 1014

 3061 GGAGTGCTCACTCAAACGCtGGAC 3084
      :::::||||| |||:::::
 1015 AlaMetLeuArgGlnIleuGlu 1022
```

```
OX   NCBI_TaxID-6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RP   Submitted (OCT-1996) to the EMBL/GenHank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE-99069613; PubMed-9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z81122; CAB54313.1;.
DR   InterPro; IPR001357; BRCT.
DR   InterPro; IPR002965; P_rich_extensn.
DR   Pfam; PF00533; BRCT; 4.
DR   PRINTS; PR01217; PRICHEXTENSN.
DR   SMART; SM00292; BRCT; 4.
DR   PROSITE; PS50172; BRCT; 4.
SQ   SEQUENCE   1076 AA;  120541 MW;  206A94A192802843 CRC64;
```

alignment_scores:
```
          Quality:   524.00          Length:        692
            Ratio:    1.444            Gaps:         24
Percent Similarity:  52.457   Percent Identity:     25.000
```

alignment_block:
US-09-664-641-10 x Q9U370

Align seq 1/1 to: Q9U370   from: 1 to: 1076

```
1350 ACAGACTCTTTCAGCCACC..........TACGAAAAATTAGAACAGCAGGTGA 1393
     ||||    ||||||||||:::           :::::|||||||||||| |  :::
 349 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 365

1394 ATCACAGCCAGCAGGGACATACAAAATGCCAATGCAGTCGTCTGTTTAGCCAA 1443
     ::|||||  |   |||  ||  |||||||||||  ::||||   |||| ::
 365 oThrAlaProProGlyMetArgGln..........TyrIleSerProG 378

1444 GTTGAAAGTGACTCCAGAGAGACAGACACATGCTACACCA..........GCCAGCA 1484
     |||  |||   | :  :||||||||||||||||| |    ||          |||
 378 lnAlaGlyGlnGlnGlnProThrProGlyThrProHisArgPheAlaPro 394

1485 CCAGCCCACCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
     ||||||||:::|||||||||||  ::|||||::|||::|||||||
 395 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 411

1535 AGCAGATAAT.......................GCAGCTCCAGCACCAC:AGCAG 1566
     |||||| :||                       ::|||  |||||||||||||
 411 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnSer 427

1567 CACCAG..........................CCGCCGCAT...  1608
     ||||||                          |||||||:::
 428 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 444

1573 .......................ATCTCTCAGCAGAACTTACCCCCAGCAG. 1599
     .......................||||||||||    |||||||
 444 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 461

1600 .......................CCGCCGCAT...  1608
                            |||||||:::
 461 InGlyThrProGlnProIleLeuSerAlaSerArgSerLeuProProGlnSer 477

1609 .......CCATTTTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCCACC 1646
        |||||||||||  |||  |||||||||||||||||:::
 478 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 494

1647 ACCATCGCCTCAGCAGCACATCAGCTTTTGGACATGATCCA. 1686
     |||||||||   ||||||  |||||||
 494 nThrMetAsnGlnMetHis..HisAspValAsnProPhep 510

1687 .....GCAGTGGAGATTCCAGAAGGAGAAGGCTTCTATTGGGATGTGTGTTT 1731
```

(right column)

```
 510 rolIleGlnMetAsnValAspProGluLeuTyrLeuThrGlyIleHisPhe 526
     :::::|||||
1732 GCAATTGCGGATTATCCAGAGCAGCAGATGTCTGATAAGCAACTGCGCCAC 1781
     :::  ||| |||||||||||| |||||||
 527 HisCysPheAspSerGluLysSerLeuPheGluAspArgLysAs 543
     :
1782 CTGGAAAAGGATAATCCAGGCACATGGGCGGCGACTGTTGAC..........C 1822
     ||| :|||||  |||||||||||||||||| :::
 543 nLeuGluPheMetIleLeuTyrHisGlyIleAspIleGluAlas 560
     ::|||
1823 CCACCTTCACGAGTCGA..........TGCACGCACCTTCTCTGTGAGAGT 1863
     |||||||||||      ||||||||
 560 erLysPheHisGluArgIleHisGlnIleThrHisValLeuAspSer 576
     :::|||
1864 CAAGTCAGCAGCGCGTATGCACAGGCCAAATAAGACAAAAAGACAAGTGT 1913
     ::::::  ||||||||||||||||||||||:
 577 CysArgGlnAsnThrAsnValMetThrSerLeuGluHisArgLysArgIleVa 593
     :::|||||
1914 TACTCCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAAAAATGGTACCGC 1963
     |::||||||  |||||||||||||||:::::|
 593 lSerIleGlnTrpIleLeuGluSerProValMetGluArgGlyValGluLeup 610
     :::
1964 CGCACGGAGCCCTTCACTCCCCAGTGGCCTTCCCACCAGGAGGAAAGCCA 2013
     ||||||| :::|||:: |  |||||||::||||||||| :::
 610 roTrpArgIleAlaHisLeuProProProPheAsnGluAsnPheArgPro 626
     :::|||
2014 TGTTCACCATATTATTTCTGTGACTGGATTTGTTGATTGTGACAGAGA 2063
     :::::|||||||||||||||||:::||||
 627 TyrLeuGlyLysLeuPheSerLeuSerGlyPheAspSerSerGluArgAl 643
     :::|||
2064 TGACCTAAAATAAATGGCTTTATTGGCAGGTGCCAAAATATACGGGTTATC 2113
     ::::|||||:||| |    ||||||||||:: :::|
 643 aAlaIleSerPheMetAlaGluMetMetGlyAlaLysIleThrProPheL 660
     :: :||||
2114 TATGCCGCAGCAACACAGTCCTCATCTGTAAAGAACAACTGGTTTAAAG 2163
     ||||:::  ||||||||||:::::|||||
 660 euAlaArgGlnAsnAsnProPheAsnGluLeuLysGlnProAlaAsp..Lys 675
     :||
2164 TATGAAAAAAGCCAAAGCAGTGGAGGATCAACCCGTGTGCACGCCCAGTGGCT 2213
     ::::|||||||::|||||   ||:||||||||| ::
 676 ValThrArgAlaGlnGluTrpLysValProValAlaAspPheGlnTrpIl 692
     :::||||
2214 TGGGCACAATTCTTCTCGGGAAACTTTGAGGCACTGAGGCAG.......... 2253
     :::|||||||||||||||
 692 eAlaAspAlaLysTyrArgValCysSerAsp..GluAlaSerProGlnLysAspProIla 708
     :::|||
2254 .ATTCAGTAIAGTCGTACACGGCCATTCAGTCTGCAGGATCCCATTTGCC 2301
     ::::|||:::||||||::|||
 708 snValGluAsnProValTyr.........GlnLeuGluAsnProValMet 721
     :::
2302 CCTACCCAGCATTACTTTTAAAT............................ 2325
       ::  |||
 722 ThrAsnValAsnThrSerProAsnHisIleGlnGlyIleAlaSerGlnGluPh 738
     :::|||
2326 .......CTTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGCTGCAGAGT 2368
     :::::||||||||||:: :::
 738 eGlyAlaMetSerCysTrpLysSerGlnThrIleIleAlaAspGlnA 755
     :::
2369 TGTTGAIGGAGTATAAGACTACCTCCCAAAACTGAAACAGAATGAAGTA.. 2415
     |||::||||||||||||||||:::|||
 755 laTyrGluLysGlyAlaArgGlnAsnArgLysThrLeuGlnAsnAsnThrPhe 771
     :::
2416 ........GCTAAATGTCCAGCCTTCT.......... 2433
        ::||||||||:::
 771 PhePheProSerLysArgLeuSerAspGlnAlaValAlaAlaProThrGlu 788
     :||
2434 .......TCCCAAAGAGCCAGAATGAAGACGTA. 2460
     ||||:::|||||||||||||:::
 788 pGluIleLysSerSerArgAlaArgLeuAspGluIleLysThrHisTrpG 805
     :::
2461 ...........CCACCTCCCACTAAAAAGCTA... 2481
     |||  :::|||::||||:::
```

RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RP   SEQUENCE FROM N.A.
RN   [1]
RX   MEDLINE=96281124; PubMed=8724849;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
DR   EMBL; D79992; BAA11487.1; -.
DR   InterPro; IPR001357; -.
DR   InterPro; IPR000253; BRCT.
DR   Pfam; PF00498; FHA; 1.
DR   Pfam; PF00533; BRCT; 1.
DR   SMART; SM00292; BRCT; 1.
DR   SMART; SM00240; FHA; 1.
DR   PROSITE; PS50172; BRCT; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
DR   PROSITE; PS50006; FHA_DOMAIN; 1.
SQ   SEQUENCE   2089 AA;   226690 MW;   A628ECC7567EEE50 CRC64;

alignment_scores:
         Quality: 342.50   Length: 842
         Ratio: 0.839   Gaps: 33
Percent Similarity: 48.456   Percent Identity: 22.922

alignment_block:
US-09-664-641-10 x Q14676

Align seq 1/1   to: Q14676   from: 1   to: 2089

(The body of this page consists of a rotated DNA/protein sequence alignment. The aligned sequences with position numbers and match markers are reproduced below as they appear.)

seq_name: sp_human:Q14676

seq_documentation_block:
ID   Q14676   PRELIMINARY;   PRT;   2089 AA.
AC   Q14676;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   KIAA0170 PROTEIN.
GN   KIAA0170.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RC   TISSUE=MYELOID LEUKEMIA CELLS;
RP   SEQUENCE FROM N.A.

2040 IleThrCysProGlnAspPheProHisCysSerIleProLeuArgValGl 2056
3027 CATAGAAGTTCACAATGCAGAGTTCGTTCGTCGACTGGTGCTCACTCAAA 3076
2056 yLeuProLeuLeuSerProGluPheLeuLeuThrGlyValLeuLysGlnG 2073
3077 CGGTGGACTATGAATCATAT 3096
2073 luAlaLysProGluValPhe 2079

seq_name: sp_human:Q960C2

alignment_scores:
       Quality: 340.50
       Ratio: 0.833
Percent Similarity: 48.575   Percent Identity: 22.803

alignment_block:
US-09-664-641-10 x Q960C2

Align seq 1/1 to: Q960C2 from: 1 to: 2090

Length: 842
Gaps: 33

688 ATTGTTCAGAGCCAAAGGGGGGGAAATACGAAGTGTGCTTAAGCGAGC 737
     :::::::::: :::::: |||||||||:::: ||:|: ::::|||:
1384 ValThrProGluProThrSerArgAlaThrArgGlyArgLysAsnArgSe 1400

738 AAGTATTAAAATTCTGACTCCTGACTCGGTTCTCGGAATTGCCTATCAGAGA 787
     ::::::||| ||| |||||:::   ::   :::
1400 rSerGlyLys........ThrProGluThrLeuValProThrAla 1412

788 AAACCAAAAAGGACGAAGCATTTTATCATCCTGGTCTCTGGATTATTTATGAA 837
     |||:|:||| ::::::: :::
1413 ....ProLysLeu.............. 1415

838 GAGGAAGAAGAAGAAGAAGAAGAAGGAAGAAGAAAATGAGGAACA 887
     ::: :::   ::::::::: |||||||||:::: |||
1416 .....GluProSerThrSerArgSplnProValThrProGluProTh 1430

888 AGATTCTCAGAATGAGGGTAGTACAGATGAGAAGTCA.................. 924
     :::|:| |||||||||:::::::::::
1430 rSerGlnAlaThrArgGlyArgThrAsnArgSerSerValLysThrProG 1447

925 ...........AGCCCTGCCAGCTCTCAAGAAGGGTTCCCTTCAGGTGAC 963
     ||| ||| ||| |||:::
1447 luThrValValProThrAlaProGluLeuGlnProSerThrSerThrArgp 1463

964 CAGGAGTTTTCACCTAATCCAACACTGAAAAATTCTAAAGGGGGAATTAAT 1013
     ||| :::|||::: :::|||:: ::|||:::
1463 1474

1464 GlnProValThrProGluProThrSerGlnAlaThrArgGlyArgThr.. 1479
     ||| ::: ||||||:::
1014 GTTTGATGGATTCTTCAGATTCATCACCGGAAAAACAGGAGGAAAATTTAA 1063

1480 ...AspArgSerSerValLysThrProGlu.............. 1489
     ||| :::: |||||| ||
1064 ACTGGACACCCCGGCCGAAGTCCCACAGTTA....GCTGCAGCAAAAGGCAGG 1110

1490 .....ValValProThrAlaProGluLeuGlnAlaSerAlaSerThrAsp 1504
     :::|||::: |||||| ::: ::: :::
1111 CTGCCTCAGGGAAAGGAGCCT............................GGGTTGATTAA 1142

1505 GlnProValThrSerAspProThrSerArgThrThrArgGlyArgLysAs 1521
     :::|||:::||||||:::::::::
1143 CTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGG 1192

1521 n............................ArgSerSerValLysThrProGlu. 1529
     |||||| ||||||:::
1193 TCCGGGGGTGTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGA 1242

1530 ....ThrValValProAlaAlaProGluLeuGlnProProThrSer 1543
     :::|||::: |||::: :::
1243 TCAGAAATGATAGCTACCTGCAGTCCCAGCTGTGACCGACACTGAGGAAATAT 1292

1544 ThrAspGlnProValThrProGluProThrSerArgThrAlaThrArgGlyAr 1560
     ::: |||::: ||| :::|||:::
1293 TACTAATAATGCTGACATTCAGACAGATGAACCGGCCATCAAAATGTAGCAC 1342

1560 gThrAsnArgSerSerValLys.........ThrProGluSerIleValP 1574
     |||||||::::::::: |||::::|||
1343 ATATCTTACAGACTCTTTCAGCCACCTACAGAAAAATTTAGAA.......... 1383

1574 roIleAlaProGluLeuGlnProSerThrSerArgAsnGlnLeuValThr 1590
     ||| ||||||:::   :::::: ||| :::
1384 ...CAGCAGGTGAATCACCAGCCAGGAGACATACAAAATGCCAATGCAGT 1430

1591 ProGluProThrSerArgAlaThrArgCysArgThrAsnArgSerSerVa 1607
     :::|||::: :::||||| ||||||:::
1431 GCTGTTTACCTTCAGCCCCAGCAGAGATAATGCGGCTCCAGCAGCAGCA 1474

1607 LlysThrProGluProThrValProThrAlaProGluProHisProThr. 1623
     ::: |||:::|||
1475 AGCAGCAGCAGCAGGGCCCAGCAGC.........AGCAGCAGCAGCACCCG 1515

1624 ThrSerThrAspProGluProValThrProLysLysThrLeuThrArgAr 1640
     ||| :::|||:::::: :::|||||
1516 GTTTTACACCTTCAGCCCCAGCAGCAGAGATAATCGGCCTCCAGCAGCAGCA 1565

1640 gArgLysThrAsnArgSerSerValLysThrProLysProValGluPro 1657
     ||| ::: |||::: ||||||:::
1566 GCAGCAGCATCCTCAGCAAACCTTACCCCAGCAGCCGCCGGCCGCATCCAGCA 1615

1657 LeuAlaSerAspLeuGluProGluProThrProAspProArgSerGlnV 1673
     :::|||:||:::||||||:::
1616 CACAGCAGCAGCAGCAGCGAACAGCACCACCACCTCCAGCAGCAT 1665

1673 ThrProGluThrProLysSerThrLeuThrArgSerArgThrVa 1690
     ||| ::: :::::: |||:::
1666 GGuAlaIleAlaGlnGlyGlnSerLysThrLeuGlySerThrAsp 1690

1690 LArgAlaMetProValProThrProPheGlnSerGluProValThrT 1707
     ||| ::: ||||||::: :::
1674 GluAlaIleAlaGlnGlyGlnSerLysThrLeuGlySerThrVa 1690

1707 hrAspGlnProIleSerProGluProThrGlnProSerCysIleLys 1723
     :::||| |||::: |||||||
1690 lArgAlaMetProValProThrProPheGlnSerGluProValThrT 1707

1724 ArgGln.........ArgAlaAlaGly.AsnProGlySerLeuAlaAla. 1736
     ||| |||:::|||||:::
1766 AGCAACTGCTGGCCACCTGGAAAAGAATAATCCAGGCACAGTGGCGGCACT 1815

1816 GTTGACCCCACCTTCACGACTGCGATGCACCGGCACCTTCTCGTGTGAGCA 1865
     ||| :::::|||
1751 ProIleAspHisLysProCysSerLeuAlaProLeuGluProLysSer 1751

```
1866 AGTCAGCACGCGTATGCACAGGAGACGAATAAGAGAGAAAGAGATGTGTTA 1915
          ::: |||||::|||
1752 GlnAlaSerArg.........AsnGlnArg...................... 1758

1916 CTGCACACTGGTTAAACACAGTCTTTAAAAAAAAAAAAATGTACCGCCG 1965
          :::||||       |||||||
1759 .........TrpGlyAlaLavAlaArgAlaAlaGluSerLeuThrAlai 1771

1966 CACCGAGACCCTTCACTTCCCAGTGGCCTTCCCACC.............. 2000
          ........ |::|::||| |||| |||
1771 leProGluProAlaSerProGlnLeuLeuGlyThrProIleHisAlaSer 1787

2001 .................AGGAGCAAACCCATGTTCACAGCATATTA 2029
          :::: |||||||| |||
1788 GlnIleGlnLysValGluProAlaGlyValArgSerArgPheThrPro... 1802

2030 TTTCTGCTGCGATTTGTTGATACTGACCCAGAATGACCTAAAATTAATG 2079
          :||||||  |||||| :::
1803 ..................GluLeuGlnProLysValSerG 1810

2080 GCTTATTTGGCAGGTGCCAAATAATACGGGTTGTATCTATGGCCGAGCAACAC 2129
          :::|| ||:::|||
1810 InSerArgLysArgSerLeuAlaThrMetAspSerProProHisGlnLys 1826

2130 AGTCCTCANTCTGTAAAGAACCAAACTGGTTTAAAGTATGAAAAAGCCAAAG 2179
          :::||||| |||
1827 GlnProGlnArgGlyGluValSerGlnLysThrValIleIleLysGluGl 1843

2180 AGTCGGAGGATACCCTGTGTCAACGCCCCAGTGGCTTGGCCACATTCTTCTG 2229
          ||||||| ::::
1843 ..................AlaGlyLysPro.............. 1851

2230 GGAAACTTTGAGGCACTGAGGCGAGATTCAGTATATAGTGCGCTACACGGCATT 2279
          ||| ||| |||
1852 ..................GlyLysGluGluAspValVal. 1858

2280 CAGTCTGCAGGATCCCATTTGCCCCTACCCAGATTTAGTGTTTAAATCTTT 2329
          |||  |||:::
1859 ..................ThrProLysProGly......... 1863

2330 TAGATCCTTGGAGAGTTCCTTAAAAGTGTCTCCACAGAGTTGTTGATGAGT 2379
          ||||||| |||||||||||
1864 ..................L 1864

2380 ATAAGACTACCTCCCAAACTGAACACNGAATGAAGTAGCTAATGTCCAGCC 2429
          ||||||||||||||||
1887 uAsnGln.GluSerThrAlaProLysValLeuPheThrGlyValValAsp 1903

2430 TTCTCCCAAAGAGCCAGAATTGCAAGAGCTACCACCTCCCACTAAAAAGC 2479
          :::|| |||||||
1876 ArgIlePro...SerArgSerLeuArgArgThr............. 1887

2480 TAACTCCACGAATTGACC......CCTTTTGTGCCTTTCACTGCATTCGAGCCT 2526
          |||||||||||||||||
1887 uAsnGln.GluSerThrAlaProLysValLeuPheThrGlyValValAsp 1903

2527 GTCCAGGTTCAACAGTATATTAAGAAGCTTCACATTCTTGGTGGAGAGGT 2576
          ||||||||||::|:
1904 AlaArgGlyGlu........ArgAlaValLeuAlaLeuGlyGlySerLe 1917

2577 TGCGGAGTCGTCGCACAGAAGTGCACACCCATTCCAGCAAAAGTGACTC 2626
          :::||  |||||
1917 uAlaGlySerAlaAlaLaGluAlaSerIleLeuValThrAspArgIleArgA 1934

2627 GCACCGTCAAGTTCCTGACGCGGCCGATTTCTGCGTGAAGCACACATAAGTACG 2676
          |||||||||||||||||||
1934 rgThrValLysPheLeuCysAlaLeuGlyIleProIleLeuSer 1950

2677 CCAGAGACTGGGCCTGGAAGAAATAGCTTCAGGTGTGTCAGGAAGTTCAATTGAGGAGCA 2726
          ::|||||||||
1951 LeuAspTrpLeuHisGlnSerArgLysAlaGlyPhePheLeuProProAs 1967
```

```
2727 GAACTACATATCTCCCAGATTGCCTGAGGCGAGAAGTACTTTTCTCTTTCAGCT 2776
          ::|||:|||:|||| |||
1967 pGluTyrValValThrAspProGluGlnGluLysAsnPheGlyPheSerL 1984

2777 TGGAAGAATCCTTAAAACGGCGACACAGTTTCTCCACTCTTTAAGGCAAAA 2826
          |||||||||||||||||:::
1984 euGlnAspAlaLeuSerArgAlaAlaArgGluArgLeuGluGlyGlyTyr 2000

2827 TATTTTACATCACACCTGGAAATCCTGCCCCAAAGTCTTTCCACTATGAAGGC 2876
          |||:::||||||||||||:::
2001 GluIleTyrValThrProGlyValGlnProProProGlnMetGlyGl 2017

2877 AATCGTAAGAGTGTGTGCAGGAGGGAAAAGGTGTTATTCCAAGCACCAGCCATCTTTCC 2926
          |||||||||||||||||||
2017 uIleIleSerCysCysGlyGlyTyrThrTyrLeuProSerMetPro..... 2031

2927 GGAAGCTCATGGAGCACAAGCAGAGACTGAAGTTTGTCGGAAAATAATTTTA 2976
          :::|||
2032 ................ArgSerTyrLysProGlnArgValVal 2040

2977 ATAATCCTGTGAAAATGACCTTCATTTATGCCGAGAATATTTTGCCAGAGG 3026
          :::|||:::||
2041 IleThrCysProGlnAspPheProHisCysSerGlnSerIleProLeuArgValGl 2057

3027 CATAGAGTGTTCACAATGCAGAGTTCGTTCTGCACTCGAGTGCTCACTCAAA 3076
          |:::||       |||||||||||||| |||
2057 yLeuProLeuLeuSerProGluPheLeuGlnThrArgValValLeuLysGlnG 2074

3077 CGCTGGACTATGAATCATAT 3096
          ||||||||
2074 luAlaLysProGluAlaPhe 2080
```

alignment block:
US-09-664-641-10 x Q9LVF7

Align seq 1/1 to: Q9LVF7 from: 1 to: 1041

183 GCACAGGTTATTCAGCTTCTCAAGGCTGAAAA..................GCGAAGGAAGT 226
    :::: ||||||||||||||||||||||:::: ||||||||:||
330 SerLysPheValGlnIlePheLysGlyProGlnLeuAlaLysLysVa 346

227 ..............AATGCACTAGCCTCACACATAATCTCAGGAGGATGGGG 270
                  ||||||||||:: || ::||||||||||:||
346 lSerTyrLysSerGlyValAlaValGlyAsnAsnIlePheAspIlePAspAspA 363

271 ACAATCCACAGAGTGGGA................................ 287
    :::: ||| |||
363 snArgGluAspGluGlyCysGlyAsplleTyrArgArgArgLysAspGlu 379

346 TTCCTTAC.............AATGCACTAGCCTCACACATAATCTCAGGAGGATGGGG 394

288 ..............GAAGCTCGGGAAGTCTTTGACTTACCTGTTGT 319

380 PhePheGlyValAlaSerLysArgArgGluPheSerSerLeuPro.... 394

320 AAAGCCTTCTTGGGTATTCGTCCGTTCAGTGTGGAACTCTTCTCCCAG 369

395 ....ArgGluGlyLysArgGluLeuIleProV 404

370 TAAAAGGTTTTCTCCAGAATCATCGATTTTTTGGAATCACTGCC 419

404 al 404

420 TGCCTTCTCAGGTGGTTGATACAAGCTGG 457

405 ....AlaValAspLysArgTrpAlaArgSerAspSerLysLe 417

458 GTTGGAGTCTTCCAGAGCTCCCGAAGACGAGAGGTAGGCAAGG 500

417 uLeuLysHisSerValThrArgSerArgValAsnIleGlnGlyAlaLysL 434

501 .AGCTTGTCCAGAGAAGAGTTGGGAAGCACAGAGAGATCATCTGCCTTCTC 548

434 ysAsnLeuGlyLys.GlnLeuAspGluValArgGluAlaValAlaValLeuG 450

549 TGACCCGGTATTGATTGCAGCCTGAGCCTCTGTGTGTAATGTCCTGGGTGT 598

450 yAsnAspThrGlnIleValAlaAlaAlaGluAlaLeuAlaCys.......... 464

599 CATCTGAAGACAGGAAGTCCCCTCTGGCCTTTGGTTACGTTCTATGGGGA 648

465 .SerGlyAspAlaArgGlyLys............PheAspGly 473

649 GATTGCCAGCTACCCTCAATAAGAAATGCACGCATTTGATTGT.CCAGA 698

474 GluAlaSerCysLeuThrGlyLysLys.............LeuSerProGl 486

699 GCCAAAGGGGGGGAAAAACGAANTGGCTTAAAGCGAGCAAGTAATAAAA 748

486 GCCAAAGGGGGGGAAAAACGAANTGGCTTAAAGCGAGCAAGTAATAAAA 490

486 uGluGluArgGly......... 490

749 TTGTGACTCCTGACTGGCTTCGGATTGGTATCAGAGAAAACCAAAAG 798

491 .PheSerProGlyValThrArgGlnSerLysGlyThrGluArg 506

799 GACGAAGCATTTAATCACTCCGTCTGATTATTATGAGAGGAGAGA 848

507 IleGlnAlaMet............. 510

849 GGAAGAGGAAGAGGAGGAGTAGAAAAATGAGGAACAAGAGATTCTCAGA 898

511 .SerLysAspGluGluLeuGlyLysArgMet............ 520

899 ATGAGGGTAGTACAGAGATCAAGCC....TCCAGCTCTCAA 942

521 ....LysLysAlaSerProSerProAlaLysAlaCys 531

---

943 ...........GAAGGGTCTCCTTCAGTGACCAG............966

532 ArgThrAsnIleGluGlySerSerAsnGlyAspGlnLeuAsnLysGluG 548

966 ....YProCysCysTrpLysSerArgLysValGlnThrAlaSerArgGluThrL 566

548 ...........CAGTTTCACCTAAATCCAACACT 990

565 ysLysAsnLeuValAspGluPheAspGluValSerGlnGluSerAsnThr 581

991 GAAAAATCTAAAGGGGAAGAATTAATGTTTGAT......GATTCTTCAGATTC 1034

603 etAsnAlaLeuHisSer...........GlyAspGlyArgGlu 613

1035 ATCACCGGAAAAACAGGAGAGAAAATTAAACTGGAGCCCCGCCCAAAGTCC 1084

592 aGlyProAspThrGln...........MetAlaAlaGluValM 603

1085 CACAGTTAGGTGCAGCAAAACGCAGGCTCCCTCAGGGGAAAAGGACCCTGGG 1134

603 ....GlyAspGlyArgGlu 613

1135 TTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCC 1184

614 .IleLeuAspProGluProAsnAsnLeuLeuIleGl 623

1185 CCCTGAG......GTCCGGGGTAATTTAATGGCTGCTGACAAACCTC 1228

623 yLysLysLeuLeuLeuLeuGlyLysGlyIleLeuSerArgCysGlyValValThrA 640

1229 AAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACGG 1278

640 rgLysSerLysArg...........IleLys 646

1279 ACACGTGAGGAAATATTACTAATAATGCTGACATTCAGCAGTGAACCGGCC 1328

647 GlyIleGlnAlaValAspAsn...........LysPro 660

1329 ATCAAATGTAGGCACATATCTTACAGACTCTTTCACCACCTAGGAAAAATT 1378

660 oLysAsn..........LysLysAlaArgSerIleLeuAlaLysSerP 673

1379 TAGAACAGCAGTGAATCACAGCCCAGCAGGGACATACAAAATGCCAATGCA 1428

673 heGluLysAsnMetAspArgTyrSerLysAsnAsp.......... 684

1429 GTGCTGTTTAGCCAAGTGAAAGTG...ACTCCAGAGACACACATGCTACA 1475

685 ....LysValAspThrProAlaValAlaValAlaSe 695

1476 GCAGCAGCAGGCCAG...CAGCAGCAGCACCACCAGCAGCCGGTTTAC 1522

695 rThrThrGlyLysArgGlnGlyLeuSerAsnLysHisCysMetSerL 712

1523 ACCTTCACCCCAGCAGATAAATGCAGCTCCAGCAGCCAGCGAGCAG 1572

712 ysLeu.............LeuLysGlnSerHisArgGlyGlu 721

1573 ATCTCTCAGCAACCTTACCCCAGCAGCGCCGCGCGCCGCCATTTTCACAGCA 1622

722 AlaGluValLeuAsnTyrProLysArgArgSerAlaArgIleSerGl 738

1623 GCAGCAGCAGCAACAGCAGCCACCACCATCGCCTCAGCCAGCAGCTTT 1672

738 nAspGlnValAsnGluAlaGlyArgSerSer........... 748

1673 TTGGACATGATCAGCAGTGGAGATTCCAAGAAGGCTTCTTATTGGGCA 1722

749 ....AspProAlaPheAspThrProAlaLysSerThrProSer 762

1723 TGTGTGTTTG..AATTGCGGATTATCCGAGCAGCAGATGTCTGATAAGCAACT 1772

```
763 ThrAsnValSer...........................ProIleCysMetGlyAsp..........  772
                                            |||::::|||
1773 GCTGGCCACCTGGAAAAGGATAATCCAGGACACATGGCGGCACTGTTGACC                1822
                                                      ::::
773 .........................GluTyrHisArgLeuSerCysLysA.                   781
                                           |||
1823 CCACCTTCACAGATCCATCGACCACCACCTTCTCTGTGAAGTCAAGTCAGC                1872
     ::::||||||||||||::|::  |||
781 spSerPheThrSerHisThrThr..............................                 788
                              |||
1873 AGCCGGTATGCACAGGCAATAACGAGAAAGAAAGAGATGTGTTACTCCACA                1922
                                                        ||||||
789 ..........ArgGlu.....................................                 790
                    |||||
1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAATGTACCGCGCACCGAG        1972
                                                  ||||:
791 ..........................................PheArgS      793
                                               |||:
1973 CCCTTCACTTCCCACGTGGGCTTCCCACCAGGAGAAACCCATGTTCACAG    2022
     :::|||  ||||||||| |||
793 erLeuThrValProValAlaGluPro.........................   801
             |||||||||||||:::
2023 CATATTATTTCTGCTGACTGGATTTTGTTGATAGTGCACAGAGATGACCTAAA 2072
                                   |||||
802 .................................LeuGlySerIleCys.....  813
                                     ::::|||
2073 ATTAATGGCTTATTTGTCAGGTGCCAAAATATACGGGTTATCTATGCCGGCA  2122
                                                   |||
814 .....................................LeuGlySerIleCys.. 818
                                         |||
2123 GCAACACAGTCCTCATCGTGTAAAAGAACCAACTCGTTAAAGTATGGAAAAA  2172
                          ||||||||
819 .....ValLeuPhe...................................... 821
           ||||:::
2173 GCCAAAAGAGTGGACGAATACCTGTCTCAACGCGCCAGTGCCTTGGCGACAT 2222
                                         :::||
821 ..................................................  821
2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCGAGATTCAGTATAGTCGCTACA 2272
                                         |||
821 ..................................................  821
2273 CGGCCATTCAGTCTCGCAGGATCCATTGCCCCTACCCAGCAGTTAGTTTTA 2322
                                         ::::|||
822 ..............................SerGlnHisLeu.......... 825
                                  |||:::||
2323 AATCTTTTAGAATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTT 2372
                                                   |||
825 ..............................................  825
2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCCTAATG 2422
                                         :::||
825 ..............................................  825
2423 TCCAGCCTTCTTCCCAAAAGAGCCAGAATTGAAGACGTACCACCTCCCACT 2472
                                         ||||||||
826 .............................AspGluAspValThrLysHisGln 833
                               ::::||||||||
2473 AAAAAGCTAAACTCCAGAATTGCACCCCTTTTGTGCTTTTTCACTGGATTCGA 2522
     ||||||:::
834 LysLysIle.......................LeuAlaArgPheAspIle.... 842
     ||||:::  |||
2523 GCCTGTCCAGGTTCAACAGTTAATTAAGAAGCTCTACAATTCTTGGTGGGAG 2572
                                         ::::||||
843 ...............................................S      843
                                                   :
2573 AGGTTGCGGGAGTGTCGCACACAAGTCGCACACAACCTCATTGCCAGCAAAGTG 2622
     ::  ||||:::||    ::::||::||:::
843 erGluAlaSerSerMetLysGluAlaThrHisPheIleAlaAspAsnPhe     859
     ||:::||||:::||||::::||||::||||
2623 ACTCGCACGGTGAAGTTCCTGACGGCGGCAATTCTGTCGTGAAGCACATAGT 2672
     |||||||||        ::::||
```

```
860 ThrArgThrArgAsnMetLeuGluValThrLeuIleAlaSerGlyLysProValVa 876
2673 GACGCCAGAGTGGCGTGGAAGAAGAATGCTTCAGGTGTCAGAAGTTCAGTTGATG 2722
     |||:  ::::||||||||::::    |||
876 lThrThrGlnTrpLeuGluSerIleAspAsnValAsnIleTyrValAspG       893
     ::::|||||||||:::  ::::||  |||
2723 AGCAGAAACTACACATTCTCCGAGATCCTGAGGCAGAGCAGAGAGTACTTTTCTCTTTC 2772
     ||:::  ||||||||||||||||||::  :::|||
893 luAspMetTyrIleLeuArgAspSerLysLysGlyGluPheCysSPhe         909
     ||||||:::    ::::||||
2773 AGCTTGGAAGAAATCCTTAAAACGGGACCACGTTTCCCACTCTTTAAGGC       2822
     ::::|||||||||:::  ||||||:::
910 AsnMetGlyValSerLeuAlaArgAlaArgGlnPheProLeuLeuGlnGl        926
     |||||||:::  ::::||||
2823 AAAAATATTTTTACATCACACACCTGGAATCTGCCCAGTCTTTCCACTATGA     2872
     ::::|||||:::  ||||||:::
926 yArgArgValPheIleThrProAsnThrLysProAlaLeuAsnThrIleT       943
     |||||||:::  ::::||||
2873 AGGCAAATGCTAGAGTGTGCAGAGGAGGAAAGGTGTTATCCAAGCACCAATCT 2922
     ::::||||||::::  |||
943 hrThrLeuValLysAlaValHisGlyLeuPro.....................  953
     ||||||:::  ::::||
2923 TTCCGGAAGCTCATGGAGACAAGACAAGAACTCGACTTTGTCGGAA         2967
     ::::|||||||:::  ||||
954 .................ValGluArgLeuGlyArgSerSerLeuSerGluAspGly 966
     ::::||||||:::  ||||
2968 ..........ATAATTTTAAATCCTGTGAAAAATGACCTTCATTTAT        3004
     ::::||||||:::  ||||
966 sValProGluAsnLeuLeuValLeuSerCysGluGluAspArgAlaIleC      983
     ||||||:::  ::::||
3005 GCCGAGAATATTTTGCCAGAGGCCATAGCAGTTCACAAATGCCAGAGTTCGTT 3054
     ::::||  |||
983 ysIleProPheLeuGluArgGlyValAlaGluValTyrSerSerGluLeuLeu    999
     ::::||||||:::
3055 CTGACTGGAGTGCTCACTCAAACGCCTGGAATCTATCACTATATAAG         3099
     |||:::  ::::||||||:::
1000 LeuAsnGlyIleValThrGlnArgLeuGluTyrGluArgTyrArg         1014
```

```
622  TCGGCCTTTGCTTACGTTCTAGGGCACAGTGC.......CAGCTAACCCT  665
       ||||||||                         ||||||
  3  ...............TrpAlaLys...       HisGlyGlyLeuTyrMetGlyGlnLeuLysMe  16

666  CAATAAGAAATGCACGCATTTGATTGTTCCAGCGCAAAGGGGGGGAAAT  715
       :||| ::||||||||||||||||||||||||||||||  ||||
 16  tasn...GluCysThrHisLeuIleValGlnLeuProLysGlyGlnLysT  32

716  ACGGAAGTGCTTTAAGCGAGCAAGTATTAATTAAAATTGACTCCTGACTGG  765
       ||||||||||                   :::|||
 32  yrGluCysAla..ArgArgTrpAsnValHisCysValThrLeuGlnTrp  47

766  GTTCTCGCATTGCCTATCAGAGACAAAAAAGGACGAAGCCATTTTATCA  815
       |||                             ||||||||
 48  PheHisAspSerIleGluLysGlyPheCysGlnAspGluSerIleTyrLy  64

816  TCCTCGTCTGATTATTATTGAAGAAGAAGAAGAGAAGAAGAAGAGAGAGG  865
       :||||||||||||||||||||||||||||||||||||||||||
 64  s..............AlaGluThrArgValGluAlaLysMetValP  75

866  AGGAAGTAGAAAAATGAGGAACAGATTCTCAGAAATGAGGGTAGTACAGAT  915
       |||||:                :|||||
 75  roAspThrSerThrProThrAlaGlnSerAlaAsnAlaGluSerHisThrLeu  91

916  GAGAAGTCAAGCCCTCCCAGCTCTCAAGAAGGGTGTCCTTCAGGTCACCA  965
       |||||||:                |||
 92  AlaAspValSerHisIleIleSerAsnIleAsnGlySerCysIleAsnGluTh  108

966  GCAGTTTTCACCTAAATCAACACTGAAAAAATCTAAAGGGAATTAAATGT  1015
       :::|||                  :|:|||
108  rMetPheGly......SerThrThrSerLysLeuGluCysSerLeu.  121

1016 TTGATGAATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAAATTTAAAC  1065
       ::|||||                |||||:|
122  ...............GluAsnLeuGluAsnLeuAsp.  128

1066 TGGACACCCGCCCCGAAGTCCCACAGTTAGCTGCACCAAAAACGCAGGCTGCC  1115
       ::||||                  |||
129  IleSerMetPheGlnAlaProGlu........AspLeuIle  139

1116 TCAGGGAAAGGAGAGCCTGGGTTGGGATTAACTTGTGTTGCCAAGTCCCACCCG  1165
       :::|||:|                |||
139  rMetPheGly..........SerThrThrSerLysLeuGluCysSerLeu  148

1166 TCCCAGGTAACATTTTGCCCCCTGAGGTTCGGGGTTAATTTAATGGCTGCT  1215
       :::||||||                |||||:
149  ...............PheSer  150

1216 GGACAAAAACCTCCAAAGTTCTCAAAGACATCAGAAATGATTAGCTTACCTGGAG  1265
       |||                    |||::|||
151  GlyArgLysLeuAspLys........  156

1266 TCCAGCTGTGACGACACTGAGAGAATATTACTAATTAATGCTGACACTTCAGC  1315
       ||||||                  |||
157  ...............LeuArgArgLeuIleAsnSerGlyGlyIleValArgPheAsnG  171

1316 AGATGAACCGGCCATCAAAAATTAGGACATATATCTTACGACTCTTTCAGCA  1365
       ||:|:||                 :|||:
171  nLeuAsn...............GluAspValThrHisValIle.  180

1366 CCTACCAAAAAATTTAGAACACACGTGAATCAC.......AGCCAGCAGG  1409
       ::|                     |||
181  .ValGlyAspTyrAspAspValArgGlnPheTrpSerLysSerSe  196

1410 ACAT.................ACAAATGCCATGCAGTGCTG.......TTTAGCC  1441
       :|/|                    |||
196  rHisArgProHisValValGlyValLysTrpLeuLeuGluCysPheThrL  213

1442 AAGTGAAAGTGACTCCAGAAGACACACATGCTACACGAGCCAGCAGCAG....  1488
       ::                    |||
213  ysGlyTyrIleLeuProGluGluSerTyrIleHisThrAsnTyrGlnPro  229

1489 ..................GCCCAGCAGCAGCAGCAGCAGCAGCACCGGTTTTACA  1523
       ::                     |||
```

```
230  AlaGlyIleAlaValSerAspGlnProGlyAsnGlnIleLeuAlaValLeuAs  246
       :::::||                  ||||||
1524 C........................CTTCAGCGCCCCAGCAGATAAATGC  1546
                                       :::|
246  pLysSerGlySerPheSerLysSerAlaLeuGluValProAlaGlnLeuG  263

1547 AGCTCCAGCAGCAGCAG.........CAGCAGCAGATCTCTGCAGCAAACCTTACCCCCAGCAGCGCCCCA  1607
       |:                       |||
263  lnGlnAlaAspGluAspProLeuLeuAlaGlnThrGlyAlaAsnAspSerThr  279

1608 TCCATTTTCACACAGCAGCAG............................CAGC  1633
       |:|| |||                 |||
280  MetValGluAlaLysSerGluValAlaLeuGluProGluValGlyProCy  296

1634 AACAGCAGCCACCACCACCATCGCCTCAGCAGCAGCTTTTTGGACATGAT  1683
       |||:|||                 :|:|||
296  sGlyProGlySerAlaHisSerLeuGluProCysAspAspSerThrHisIleSerV  313

1684 CCAGCAGTGGAGATTCCAGAGACAAGGCTTCTTATTGGGATGCTGTTGTTGC  1733
       |||||||                 :|||
313  alGlnGlnLeuHisAsnLysSerSerValSerHisCysIleLeuAspAspSer  329

1734 AATTCGGGATTATCCAGAGAGAGTGTCTGATAAGCAACTGCTGGCCACCT  1783
       |||||||                 :|:|||
330  ...............ThrValArgGluGluGlyLeuPheSerGlnLysSerPhe  343

1784 GGAAAAGGATAATTCAGGCACACATGGCGCCGCTGTTGACCCACCTTCACG  1833
       ||||||||                 :|||
343  uValLeuGlyPhe......SerValGluAsnLysCysAsnIleValAsp.  357

1834 AGTCGATGCACGCAC..................CTTTCTTTCTGTGAGAG  1862
       |||||||                  :|:|||
358  ...............IleLeuArgGluHisAlaGlyLysIle...ValSerLeuPro  370

1863 TCAAGTCAGCAGCGCGGTTATGCCACAGGCAATAAGAGAAAAACAGAGATGTG  1912
       :::|||||                 :|||
371  SerArgIleValAlaAspTyrAlaValValProLeuLeuCysGlyCysGluVa  387

1913 TTACTGCACACTGGTTAAACACAGTTCTTAAAAAAAAAAATGGTA...  1959
       |||||                   :||:||
387  lGluValThrValGly........GluValV  395

1960 CCGCCGCACCGAGCCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGGAAA  2009
       |||                    ||:|||
395  alThrAsnThrTrpLeuValThrCysIleAspAsnGlnThrLeuValAsp  411

2010 GCCATGGTTCACAGCATATATATTTCTGTGACGGATTTGTTGATAGTGACA  2059
       |||||||                 :|||:|
412  ProLysSerArgAsnProLeuPheThrProValSerAlaValMetSerGlyValTh  428

2060 GAGATGACCTAAAAATTAATGGGTTTATTTGGCAGTGGCAAATATACGGGT  2109
       ||||||||                 :|||
428  rProLeuGluAspCysValIleSerPheSerGlnCysValGlyAlaGlyGluA  445

2110 TATCTATCCCGC...................AGCAACAGT  2132
       |||||                    |||
445  rgAspSerLeuValPheLeuAlaAsnHisLeuGlyAlaSerValGlnGlu  461

2133 CCTCACCTGTAAAGAACCAACTCGGTTTAAAGTATGAAAAAGGCAAAGAGT  2182
       :|||||                  |||:||
462  PhePheValArgLysAlaAsnAlaLysLysGlyMetLeuAlaSerThrHi  478

2183 GGAGGATACCCTGTGTCAACGCCCAGTGGCTT.........  2214
       ||                     |||
478  sLeuIleValLysGluProThrGlySerLysTyrGluValAlaLysLysLysT  495

2215 ..................GGGGACATTCTTCTGGAAACTTTT......  2238
       ::                     |||
495  rpSerLeuProAlaValAsnIleSerTrpLeuLeuGluThrAlaArgIle  511
```

```
512 GlyLysArgGlnAspGluAsnHisPheLeuValAspAsnAlaProLysGl 528

528 nGluGlnValLeuGluThrLysIleProAsnGlyValSerSerAsnProA 545
2239 .........................GAGGCA 2244

545 spLeuProLeuHisProAspAlaHisLeuGluIleHisArgLysLysAla 561
2245 CTGAGGCAGATTCAGTATAGTCGCTACACGGCATTCACTCTG....... 2286

562 ValThrProLeuAspMetAsnArgPheGlnSerArgArgAlaPheArgAlaVa 578
2287 .............................CAGGATCCATTT 2298

578 lIleSerGlnGlnArgGlyGlnAspProThrPhePheProProValArgGlnP 595
2299 .......GCCCCTACCCACCCATTTA.............. 2316

595 roLeuThrLysGluProSerLeuHisLeuAspThrProSerLysProPheLeu 611
2317 ..................GTTTTAAATCTTTTAGA 2333

612 SerLysAspLysLeuPheLysProSerPheAspSerProValThrArgAlaLeuLeu 628
2334 TGCTTGGAGAGTTCCCTTAAAAGTGTTCTCGCAGAGTGTTGATGAGTATAA 2383

628 aAlaLeuGluThrProAsnAlaAlaSerGlnLysGlyArgLysLeuSerProP 645
2384 GACTACCTCCCAAACTGAAACAG 2412

645 roLeuSerGluValIleValIleValAsnLeuThrValAlaLeuAlaAsnSer 661
2413 GTAGCTAAATGTCCAGCCTTCTTCCCAAAGCCAGAAATTGAAGACGTACC 2462

662 SerArgAsnThrAspSerHisSer.AlaSerProGlnLeuLysGlyAlaH 678
2463 ACCTC.....CCACTAAAAAGCTAAACTCCAGAAT 2491

678 isLeuGluGluGluThrArgGlyTrpSerPheAsp 690
2492 TGACCCCTTTTGTGCTTTTCA.....CTGGATTCCAGCCCTGTTCCAGGTT 2535

691 ..ValValValCysValSerLysLysLeuSerLysLeuSerSerGluGlu 706
2536 CAACAGTATATATTAAGAAGCTCTACATTCTTGGTGGAGGAGGTTCGGAGTC 2585

706 uAsnGlyValAlaAlaSer.LeuGlyAlaGluTyrArgTrpSerPheAsp 722
2586 TGCACAGAAGTGCACACACTCAT.....TGCCAGAAAGTGACTCGCACCG 2632

723 GluThrValThrHisPheIleTyrGlnGlyArgLaAsnAspSerAsnAr 739
2633 TGAAGTTCCTGACGCGCGATTTCTCGCGTGAAGCACACATAGTGACGCCAGAG 2682

739 gGlu.TyrLysSerAlaLysGlyAlaGlyValHisIleValSerGluHis 755
2683 TGGGTGGGAAGAAATGCTTCAGGTGTCAGAAGTTCAATGATGAGCAGAAACTA 2732

756 TrpLeuLeuGlyCysAlaGlnGluTyrLysHisIleLeuProGluSerLeuTrY 772
2733 C 2733

772 r 772
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 84.9 KDA PROTEIN.
GN   F4C21.5 OR At4g03130.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Habermann K., de la Bastide M., Huang E.N., Gnoj L., Schulz K.,
RA   Preston R., Calma C., Martienssen R., Parnell L.D., Dedhia N.,
RA   McCombie W.R.;
RT   "Arabidopsis thaliana BAC F4C21 from chromosome IV near 17 cM.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil D.M.,
RA   Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA   Chekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA   Yewes H.W., Lemcke K., Mayer K.F.X.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RP   EU Arabidopsis sequencing project;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC005275; AAD14441.1; -.
DR   EMBL; AL161496; CAB77798.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 2.
DR   SMART; SM00292; BRCT; 1.
DR   PROSITE; PS50172; BRCT; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   765 AA;  84874 MW;  A49BB5DADE6948E0 CRC64;
```

alignment_scores:
```
          Quality:      265.50          Length:      910
          Ratio:          0.676         Gaps:          34
Percent Similarity:      43.187    Percent Identity:   19.890
```

alignment_block:
US-09-664-641-10:

Align seq 1/1 to: Q9ZR18   from: 1   to: 765

US-09-664-641-10 x Q9ZR18

```
766 GTTCTGAATTGC.........GTATCAGAGAAAACAAAAAGGACGAAGCATT 809
    ::||||   |||         ::||::|||:|::||||||:||||||:|::||
26  IleLeuLeuGlyCysGlyGluIleGluAspThrGluAlaIleAlaSerPheAsp 42

810 TTAT................CATCCTCGTCTGATTATTTTATGAAG 838
    ||||                :::||||:||:||||::||||:|:|
42  uValLeuAsnSerSerLeuLeuGluThrProPheGlnValLeuTyrAspA 59

839 AGGAAGAA.........GAGGAAGAGGAAGAGGAG 864
    :|||||||         :::|||||||||||||||
59  spThrGluGluValAspAspAsnCysGlyAspGluSerLeuAsn 75

865 GAGGAAGTAGAAAATGAGGAACAA.........GATTCTCAGAATGA 902
    |||||||::::|||||:|||||||         |:|||||:|::||
76  GluGluCysTyrGlyLeuAspGlnValValAspAspSerAspAspGl 92

903 GGGTAGTACAGATGAGAGTCAAGCCCTGCCAGCTCTCAAGAGAAGGGTCTC 952
    :|||::|||:||||||:||::|:|||:|||||||||:||||:||||||::|
92  uAspGlySerValAlaAlaGlyLeuProMetCysLeuProGluLeuSerP 109

953 CT...TCAGGTGAC 963
    ||   ::|||||||
109 roLysAlaGlyAspIleLeuLeuGluSerAspGlySerAsnAspHisGlu 125

964 ...CAGCAGTTTTCACCTAAAATCCAAC 987
    |||  ::|||||||||:|||||||||||
126 CysGlnThrGlyLysGlnGluSerAsnCysSerAspThrValThrGlyPheGlu 142
```

```
 988 ...........ACTGAAAAATCTAAAAGGGAA...TTAATGTTTG 1018
                         |||              |||:
 142 nGlySerLeuArgIleThrAlaSerSerHisGlyProGlyLeuAspPhel 159

1019 ATGATTCTTCAGATTCATCACCGAAAAACAG.......... 1050
     ||||||| |||  :::  |||  |||||||
 159 euAspSerGlnGluProGlyGluSerThrGlnAlaAspAlaLeuGlyPhe 175

1051 ..............GAGAGAAATTAAACTGACCCCGCCGAAGT 1082
                      |:||         ::|||||||
 176 ValAspHisPheLeuMetAspLysAspLeuAsnLeuSerProValAsple 192

1083 CCCACAGTTAGTTGCAGCAAAAACGCAGGCTGCCTCAGGGAAAGGAG.CTG 1132
       |||                  ||||| |:|||     ::|||
 192 uProValAsnSerGlyArgArgLysSerSerProValSerGlyValaiiysG 209

1133 GGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTG 1182
        :||||||            ||||||        ::|||||
 209 lyCysGlnSerLeuAlaLysArgIleMetSer.................... 219

1183 CCCCCTGAGGTCCGGGGGTAATTTAATGGCTGCTGGACAAAACCTCCAAAG 1232
                         :::|
 220 ...............................Argse 221

1233 TTCTGAAGAATCAGAAAAATGAATAAGCTACCTGG.............. 1263
     |||:|            :::
 221 rProThrArgLysMetSerValPheAspTrpSerSerAspGlnCysAspV 238

1264 ..................AGTCCAGCTGTACGG........ACACTGAGG 1287
                         |||||||:||:::||||   ||
 238 alSerAspProLysThrSerProValThrArgAlaSerIleThrCysSer 254

1288 AATATTACTAATAATCCT...........CACATTCAGCAGATGAACCG 1325
                                    |||  |||  :::|||  |||
 255 LysIleArgGluSpGlnValAlaGlyValAspAspPheGlyValLysLysAr 271

1326 GCCATCAAAATGTAGCACATATTCTTACAGAGTCTTTCAGCACCTACGAAAA 1375
                 |||    |:||  ||| |:|          |||
 271 gSerArgAspProCysAspIlearGlyLysValSerSerHisProThrGln 288

1376 ATTTAGTAACAGCAGGTGAATCACAGCGCAGGAGAGCAATACAAAAATGCCAAT 1425
           :::|||  :::::|||    |||
 288 rgValMetLys...............AsnSerSerAlaLys 296

1426 GCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGCACACACATGCCTACA 1475
     |||:||            ||| |||     :::|||
 297 HisLeuLysMetGluGlnAlaSerGlyLeuSerGlyIleMetPheIl 313

1476 GCAGCAGCAGCAGCCCCCAGCACNCAGCAGAGCAGCAGCACACGGTTTACC 1525
     |||||||||||||||||  |||
 313 eSerGlnlysAspAlaGlnlleuGluASpLysAla...ProLysGluProl 329

1526 TTCAGCCCCAGCAGATAATG................CAGCTCCAG 1554
     |||:||||||||||           |||||
 329 euGluProGluGluGluAspPheIleGlyIleAsnThrGlnIleAla 345

1555 CAGCAGCAGCAGCAGATCTCTCAGCAACCTACCCC. .............. 1593
        |||              :::
 346 AlaGluAlaMetSerAlaLeuValTyrAlaProCysThrThrGluGluAl 362

1594 ....CAGCAGCGCCGCATCCATTTTCACAGCAGCAGCAGCAACAGC 1639
         |||||||||||  |||  :::|||  |||
 362 aCysGluSerAspProIleProCysSerValSerGluMetArgAspGlnv 379

1640 AGCCACCACCATCGCCTCAGCAGCATCAGCTTTTGGACATGATCCAGCA 1689
     |||                |||       :::|||
 379 alSerAsnLeuSerArgArgAsnAsnAspThrIleGluGlyGluPro... 394

1690 GTGGGAGAATTCCAGAAGAAGGCTTCTTATTGGGATGTGTGTTTTCCAATTGC 1739
          |||           |||||:::|||        :::|||  |
 395 ....................GluArgAspLysLeuAsnGlyLeuPheSerAlaAl 406
```

```
1740 GGATTATCCAGAGACGAGAGATGTCTGATAAAGCAAACTGCTTGGCCCACCTGGAAAA 1789
      :::   |||:::|||         |||::
 406 aProHisArgGluArgGlnArgAspASnSerLysSerLysLysSerArgGlyLysSerPheThrAsnGluG 423

1790 GGATAATCCAGGCACACATGGGGCCGCACTCGTTGACCCACCTTCACAGATCGA 1839
      ::    ||||||||    |||  :::|||
 423 luArgThrArgAlaAsn.....................ValSerValLysThr 433

1840 TGCACGCACCTTCTCTGTTGGAGAGTCAAGTCAGCAGCCGTATGCACAGGC 1889
      |||  ||||||||    ||||||||
 433 lurgThrArgAlaAsn.........................ValSerValLysThr 446

1890 AATAAGACAGAAAGAAAGAGAGAGATGTGTTTACTGCGCACACTCGTTAAACACAGTCT 1939
          |||||||||||||||||::::|||
 446 euMetGlnArgHisHisValProProArgArgSerTrp................ 463

1940 TAAAAGAAAAAAAAAATGGTACCCGCCGCCACCGAGCCCCTTCACTTCCCACTG 1989
       ||||||||||||||:::|||
 451 CysLeuLeuAsnLeuCysGlnTrpArg.........His.ProArgAI 446

1990 GCCTTCCCACCACCAGGACGAAAAGCCATGTTCACAGCATATATTTTCTGTGCAC 2039
        |||::::|||
 463 ........................................... 463

2040 TGGATTGTTGATAGTGACAGAGATGACCTAAAATTAAIGGCTTATTTGG 2089
 463 ........................................... 463

2090 CAGGTGCCAAAATATACGGGTTATCTATGCCCGCAGCACACAGTCCTCATC 2139
      |||||||:::|||        :::|||  :::|||:|||
 464 .GlyAlaSerLeuValLysAspArgSerGluThrAsnThr.......... 476

2140 TGTAAAGAACCAAACTGGTTTAAAGTAATGAAAAAGCCAAAGAGTGGAGGAT 2189
 476 ........................................... 476

2190 ACCTGTGTCAACGCCCAGTGGGACAATTCTTCTGGGAAACTTTG 2239
      |||||                  |||:::|||  |||
 477 .LeuSerSerSerLeuAGGlySerLeus 486

2240 AGGCACTCAGGCAGATTCACTATAGTCGCCTACACCGCATTCACTCTGCAG 2289
      |||||            :::|||||
 486 erGlyThrArgGlnAla................. 491

2290 GATCCCATTTGCCCCTACCCAGCATTTAGTTTTAAATCTTTTAGATGCCTTG 2339
      :::|||
 491 ........................................... 491

2340 GAGAGTT...CCCTTAAAAGTG. ........... 2358
 505 rHisAlaSerProArgLysIleTyrAspGlySerHisGluSerProCysA 522

2359 ..................TCTCCAGAGTTGTTGATGAGTATA 2382
                         |||||||
 505 gArgAsnLeuAlaHisValArg....................... 522

2383 AGACTACCTCCCAAACTG.................. AAACA 2405
      ||| |||
 522 snLysAspPheProArgLeuPheLeuGlnLysGluLeuThrThrSerLeu 538

2406 GAATGAAGTAGCTAATGTCCAGCCTTCTTCCCAAAGAGCCAGAAATGAAG 2455
      |||||||||            |||||||
 539 GlyGlyProGlyLysIleGlyAspPheValTrpLysAspLeuArgLysAr 555

2456 ACGTACCCACCTCCCACTAAAAAAGCTAACTCCAGAAATTGACCCCTTTGTG 2505
      |||                        :::|||  |||
 555 gArgAsnLeuAlaHisValArg....................... 562

2506 CTTTTCACT...GGATTCGAGCCTGTCCAGGTTCACAGTATTATTAAGAA 2552
 563 .............................Val 563

2553 GCTCTACATTCTTGGTGTGGAGGAGGTTGCGCAGAGAGTGCACAC 2602
 564 LeuPheSerGlnAsnLeuAspProAspGluThrValLysGlnIleGlnLysLysIl 580
```

seq_name: sp_human:Q9UEB9

seq_documentation_block:
```
ID   Q9UEB9                 PRELIMINARY;     PRT;   1435 AA.
AC   Q9UEB9;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DNA TOPOISOMERASE II BINDING PROTEIN.
GN   TOPBP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Yamane K., Kawabata M., Tsuruo T.;
RT   "A DNA topoisomerase II binding protein with eight repeating regions
RT   similar to DNA repair enzymes and to a cell cycle regulator.";
RL   Eur. J. Biochem. 250:794-799(1998).
DR   EMBL; AB019397; BAA34202.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT.
DR   SMART; SM00292; BRCT; 7.
DR   PROSITE; PS50172; BRCT; 7.
KW   Isomerase.
```

alignment_block:
us-09-664-641-10 x Q9UEB9

Align seg 1/1 to: Q9UEB9   from: 1 to: 1435

SQ   SEQUENCE   1435 AA; 160550 MW; 39FF4399A076AF9B CRC64;

alignment_scores:
                Quality: 263.50   Length: 1031
                  Ratio: 0.620   Gaps: 43
Percent Similarity: 41.222   Percent Identity: 19.011

580 eMetValArgLeuGlyIleSerProAlaSerSerSerAlaAspSerThrH 597
      |||| ::: ||||||:|||||
2603 ACCTCAATTGCCAGCAAAGTGACTCCGACCACCGTGAGTTCCTGACGGATT 2652
      :: ||||||||||||| | |:|||||||: ||||
597 isPheIleAlaAspArgPheAlaArgThrArgAsnMetLeuCluAlaIle 613
      || |:||||||| ||||||| |:::||||||
2653 TCTGTCGTGAACACAATAGTCAGCCACCAGAGTGCGTGGAAGAAATGCTTCAG 2702
      ||| ||||||||||:||||::||||||||
614 AlaLeuGlyLysPheValValThrProIleTrpLeuGluSerCysAlaGl 630
      |||||||:::::|||||||||| |||||||||
2703 GTGTCAGAAGTTCATTGATTGAGCGAGAGAACTACATTCTCCGAGATGCTGAGG 2752
      |:|||| : :|||||:::|||||||||||
630 nThrArgCysLeuLeuAspGluLysSerTyrIleLeuGluArgAspIleLysL 647
      |||||| ::||||||:::::|||||||||
2753 CAGAAGTACTTTCTTCTTTCAGCGTTGGAAGAATCCTTAAAACGGGGCNCAC 2802
      | |: |||||||||||||||||:::||||
647 ysGlu....LysAspGlyPheCysLeuLeuThrSerLeuAlaArgAlaLys 662
      ::|||||||||:||||||||||||| |:|
2803 GTTTGTCCCACTCTTTAAGGCACAAAATATTTTTACATCACACCTGGAATCTG 2852
      ||||:::||||||||||||||||||
663 GlnHisProLeuLeuGlyLysPheLysValCysIleThrProSerIleLy 679
      |||:::|||||||:|||||||||
2853 CCCAAGTTCTTCCACTATGACAGGCCAATCGTAGACGTGTCGCAGAGAAAGG 2902
      ::|||| ::||||||| |||||||||||
679 sProSerArgGlyMetIleThrAspLeuValLysMetThrGlnGlyGlnV 696
      || |:||||||||||||||||
2903 TGTTATCCAAGCAGCCATCTTTCCGGAAGCTCATGGACACAAGCAGAAC 2952
      ||||||::||||||||||||||||::||
696 alValGluLaSer..........GluIleIleAlaAlaGluAspArg 708
      ::::||||||||||||:::|||
2953 TCCAGTTGTCGGAAAATAATTTTAATATACTGTGCAAAATGACCTTCATTT 3002
      ::: |||||||||||||||||||:::||
709 AsnPheProGluAspValIleuIleLeuSerCysLysLysGluAspArgAsph 725
      |||:::|||||||||||||||||||:::|||
3003 ATCCCCGAGATATATTTTTGGCCAGAGGCCATA.............. 3030
      ::||||| ::|||||||||||:::
725 eCysLeuProPheValAsnGlnGlyThrValSerSerGlnPheGlnArgL 742
      ::::||||||||:||||||||||
3031 .........GATGTTCACAATGCAGAGTTCGTTCTGCACTGGAGTGCTC 3069
      |||||||||||||||||||:::||
742 ysGlyThrLysGluIleHis..........LysPheThrAlaCysLeuLeu 756
      :::|||||||||||:||||||||
3070 ACTCAAACCGTCGACTATGAAATCATATGIAAG 3099
      ||| ||||:::|||||||||||
757 LeuSerThrThrAspIlp...........TyrLys 764
      |||::|||||:::

234 AATGCACTAGCCTCACACATAATCTCAGAGGATGGGGACAAATCCAGAGGT 283
      |||:|||||||||||||||||:|||||:||:
53 AsnValSerValThrHisLeuIleLeuAla.............GlyGluVa 64
      :::||||||||||::||::::: ||||
284 GGGAGAAGCTCGGGAAGTC.............TTTGACTTACCGTGTAA 321
      |||||| ::: |||||:||||:
64 lGlySerLysLysTyrLeuValAlaIleAlaAlaAsnLeuLysLysProIleLeuLeuL 81
      ||||||||||||||||||||||::|||
322 AGCCTTCTTGGGTGATTCTG.............TCGGTT 347
      ||||||||||||||||||| ::::
81 euProSerTrpIleLeuLysThrLeuGluTrpIleGluLysSerGlnGluLysLysIle 97
      |||||||||||::||||||||||::::
348 CAGTGTGGGAAACTCTTCTGCCCAGTAAAATGGTTTTCTCCAGAATCAIGTCA 397
      |||||||||||||||||| :::|||||||||:
98 ThrArgTyrThrAspIleAsnMetGluAspPhe..........LysCysPr 111
      ||||||||||||||::: |||
398 GATTTTTTTTGGAATCACTGCCTCGCCTTTCTCAGGCGTGTTGATACAGAGCT 447
      |||||||||||||||||:::||||||||
111 oIlePheLeuGlyCysIleIleCysValThr.............. 121
      ||||||||||||||||||:::
448 GGAGCTCTTTGTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGAA 497
      ||||||||||||||||||||::|||||||
121 ....GlyLeuCys.GlyLeu.............. 126
      ::::::||||||
498 GGGAGCTTGTCCAGCAGAAGTTGGGAAGCACACAGAGATCATCTGCCTTCTT 547
      |||||:||||||||||||||:::||||
121 ....AspArgLysGluValGlnGlnLeuThrValLysHisGlyCl 140
      :: |||||||| |||||||||||||||
548 CTGACCCGTATTGATGCAGGCTGAGGCCTTCTGCGTTGTAAIGTGCTGGGTG 597
      |||||||||||||||||||::||||||||
140 yGlnTyrMetGlyGlnLeuLysMetAsn..........GluCysThrIleLeuIleV 156
      :::|||||| ::||||||||||||
598 TCATCTGAAGACAGAGAAGTGCCCTGTGGGCTTTGGTTAGGTTCTATGGGGG 647
      |||||||||||||||||:::||||||
156 aIGlnLeuProLysGlyGlnLysTyrGluCysAla..LysArgTrpAsn 171
      ||||||||::||||||||||||:::
648 AGATTGC...........CAGCTAAACCCTCAATAAGAAAATCCACCCATTTGATTG 691
      ::::::   :::|| ||||||||||||||
172 ValHisCysValThrThrGlnThrPhePheAspSerIleGluLysGlyPh 188
      |||||||||:::||||||||::|||
692 TTCCAGAGCCAAAAGGGGGGGAAATACGAATGTTGCCTTTAAAGCGAGCAAGT 741
      |||||||||||||||||||:::||
188 eCysGlnAspGluSerIleTyr.............. 196
      ::::::::||
742 ATTAAAATTGTGACTCCTGACTGGGGTTCTGGATTGCCGTATCAGAGAAAAC 791
      |||||||||||||::||||
196 ysThrGluProArgProGluAlaLysThrMetProAsnSerSerThrPro 212
      ::| ||||||||||||::|||
792 CAAAAAGGACGAAGCACATTTTATCACCCTGCTCTGATTATTTATGAAGAGG 841
      :::|||||||::||||||||:::||||
213 ThrSerGlnIleLeuAsnThrIleAspSerArgThr........ 223
      :::|||||::||||||::
842 AAGAAGAAGCAAGAGGAAGAGGAGGAGGAAGTAGAAAAATCAGGAACAAGAT 891
      |||||||||||||||||||||||:::
892 TCTCAGAATGAGGGTAGTACAGAAGAGAGAGATCAAGCCCTGCAGCTCTCA 941
      :::||||||||||||||||||:::

```
 942 AGAAGGGTCTCTTCAGGTGACCAGCAGTTTCACGTAAATCCAACACTG 991
           |||    ||||||
 224 ........ LeuSerAspValSerAsnIleS 231

 992 AAAAAATCTAAAGGGAAATTAAATGTTTGATTGATTCATCACCG 1041
     ||||||    |||||    :::
 231 erAsnIleAsnIleSerCysValSerGluSerIleCysAsnSerLeuAsn 247

1042 CAAAAACACGACAGAAATTAAACTGGACCGGCGCCGGAAGTCCCACAGTT 1091
     |||||          ||||||
 248 SerLysLysLeuGluProThrLeuGlu........ AsnLeuGluAsnLe 260

1092 AGCTGCAGCAAAACGCAGGCTGCCTCAGGAAAGGAGGCGTGGGTTG...A 1138
     |||  |||      :::||        |||
 260 uAspValSerAlaPheGlnAlaProGluAspLeuLeuAspGlyCysGlyi 277

1139 TTAACTTCGTCGTCCCAATGTCCCACCGTTCCCAGTTACACATTTGCCCCCT 1188
     ||||:::::
 277 leTyrLeuCysGly........... 281

1189 GAGGTCCGGGGTAATTTAATGCCTGCTGGACAAAACCTCCAAAGTTCTGA 1238
           |||::|||||||||||:::
 282 ............ PheSerGlyArgLysLeuAspLys..... 289

1239 AAGATCAGAATGAATAGCTACCTGGAGTCCAGAGTTACGGACACTGAGGA 1288
     ||||        |||  |||
 290 ............ LeuArgArgLeuIleA 295

1289 ATATTACTAATTAAATGCTGACATTCACCACATGAACCGGCCATCAAATGTA 1338
     |||||||         |||  ::::
 295 snSerGlyGlyGlyValArgPheAsnGlnLeuAsn....... GluAspVal 309

1339 GCACAATATCTTACACAGACTCTTTCAGCCACCTTACGAAAAATTTAGAACAGCA 1388
     ::|||||:||:::
 310 ThrHisValIle...... ValGlyAspTyrAspAspGluLysGlnPh 324

1389 GGTGAATCACAGC......... CAGCAGGGACATACNAAATGCCAAATGCAGTGCTGT 1435
     ::||||:::           :::   |||  ||
 324 eIleHisAlaAsnTyrGlnProValGluIleProValSerHis...Gl 340

1436 TTAGCCAAGTGAAAGTGACTCCAGACAACATGCTACACCAGCAGCAG 1485
     |||||||||         ::::      ::::::|||:|
 341 ........ LeuGluCysPheSerLysGlyTyrMetLeuSerGluGluPro 354

1486 ........ CAGCCCAGCAGCAGCAGCAGCAGCAGCCGGTTTTACACCTTCA 1529
           |||:||||||||                |||   ||
 355 TyrIleHisAlaAsnTyrGlnProValGluIleProValSerHis...Gl 370

1530 GCCCCAGCAGATAAATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCTC 1579
     |||||||||:|:
 370 nProGluSerLysAlaAlaLeuLeuLeuLysLysLysAsnSerPheSerL 387

1580 AGCAACCTTACCCCCAGCACGCCCGCGAT 1608
     :::
 387 ysLysAspPheAlaProSerGluLysHisGluGlnAlaAspGluAspLeu 403

1608 ............ 1608

 404 LeuSerGlnTyrGluAsnGlySerSerThrValValGluAlaLysThrSe 420

1609 ..... CCATTTTCACAGCAGCAGCAGCAGCAA 1635
           ||||||||||            :::
 420 rGluAlaArgProPheAsnAspSerThrHisIleAlaGluProLeuAsnAspS 437

1636 ............ CAGCAGCCACCACATGCCCTCAGCAGCAATCAG 1668
                    ||||||||||||
 437 erThrHisIleSerLeuLeuGlnGluAsnGlnSerSerValSerHisCys 453

1669 CTTTTTGGACAAGATCCAGCAGTGGAG...ATTCCAGAGAGAAGGCGTTCTT 1715
     :::           |||||||||||:::
 454 Val........ ProAspValSerThrIleThrGluGlyGlyGluPh 466

1716 ATTGGGATGNGTGGTTTGCAAATTCGGGAATTATCCAGAGAGAGATGTCTCGATA 1765
```

```
  466 eSerGlnLysSerPheLeuValLeuGlyPheSerGlnGluAsnGlu.... 481
     |||              :::    ::::::
1766 ACGAACTGCTGCCCCACCTGGAAAAAGGATAAATCCAGGCAGACATTGCCGGCACT 1815
          :::::             ::|||||||:::   ||||
 482 ........ SerAsnIleAlaAsnIleIleLeuGluAsnAlaGlyLys 494

1816 GTTGACCCCACCTTCAGGAGTCGATGCACGCAC 1848
     :::|||              |||||||
 495 Ile..Met SerLeuLeuSerArgThrValAlaAlaAspThrAlaValpr 510

1849 ............ CTTCTGTGAGAGTCAAGTCAGCAGCGGCGGGTAPGCACAGGCAATAA 1894
                |||||            |||
 510 oLeuLeuGlyCysGluValGlyGluAlaThrValGly... 521

1895 GAGAAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTGCTTAAAA 1944
     ::|||||            :::     |||||         ||
 522 .......... GluValThrAsnThrTrpLeuValThrCysIleAspp 534

1945 AAAAAAAAATGGTA...CCGCCGCCACCGAGCCCTTCACTTCCCAGTGGC 1991
     |||||:::                |||  |||
 535 TyrGlnThrLeuPheAspProLysSerAsnProLeueuPhehrProValpr 551

1992 CTTCCCACCAGGAGAAAAGCCATGTTCACACGCAATAATTATTTCGTGACTG 2041
     :::                ||||||||:::   :::||||||
 551 oValMetThrGlyMetThrProLeuGluAspCysValIleSerPheSerG 568

2042 GATTTGTTGATAGTGACAGAGAGAAGATGACCTAAAATTAATGGCTTATTTGGCA 2091
     |||||         :::          ||||||::::
 568 lnCysAlaGlyLysAlaGlyGlyLysGluSerLeuThrPheLeuAlaAsnLeuLeu 584

2092 GGTGGCAAATAATACGGGGTTATCTATGCGGC 2121
     |||||:::            |||   |||
 585 GlyValaSerVa ... GlnGluTyrPheValArgLysSerAsnAlasyslysGl 601

2122 ..... AGCAACACAGTCCTCATCCTGTAGAAGAACCAACTGGTTTAAAGT 2164
           |||||||||              :::      |||||
 601 yMetPheAlaSerThrHisLeuLeuLysGluArgGlyGlySerLysSerT 618

2165 ATGAAAAAAGCCAAAGAGTGGAGGATACCCGTGTCAAACCCCAGTGGCTT 2214
     ||||||         :::                ||||||
 618 yrGluAlaAlaLysLysTrpAsnLeuProAlaValThrIleAlaIrpLeu 634

2215 .......... GGCGACATTCTTCT 2228
           :::::::       |||   |||
 635 LeuGluThrAlaArgThrGlyLysArgAlaAspSerGluSerHisPheLeuGln 651

2229 GGGAAAC....... 2235
     |||
 651 eGluAsnSerThrLysGlyUArgSerLeuGluThrGluIleThrAsnG 668

2235 ............ 2235

 668 lyIleAsnLeuAsnSerAspThrAlaGluHisProGlyThrArgLeuGln 684

2236 ............ TT 2237
                         ::::::||
 685 ThrHisArgLysThrValValThrProLeuAspMetAsnArgPheGlnSe 701

2238 TGAGGCACTGAGGCAGGATT........ CAGTAGTCGGTACACGGCCATTCA 2281
     :::|||::||||:||          ::::::|||||:
 701 rLysAlaPheArgAlaValValSerGlnHisIleAAArgGlnValAlaAlaAlaS 718

2282 GT........ 2283
     :::
 718 erProAlaValGlyGlnProLeuGlnLysGluGlnProSerLeuHisIleLeuAsp 734

2283 ............ 2283

 735 ThrProSerLysPheLeuSerLysAspLysLeuPheLysProSerPheAs 751

2284 ............ CTGCAGAGATCCATTGCCCCTACCCAGC 2311
                         |||::::||||:
```

seq_name: sp_human:Q92547

seq_documentation_block:
```
ID   Q92547          PRELIMINARY;      PRT;   1550 AA.
AC   Q92547;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   KIAA0259 PROTEIN (FRAGMENT).
GN   KIAA0259.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
DR   EMBL; D87448; BAA13389.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 8.
DR   SMART; SM00292; BRCT; 7.
DR   PROSITE; PS50172; BRCT; 7.
FT   NON_TER       1
SQ   SEQUENCE   1550 AA;  173388 MW;  86F8B66AFB6B292D CRC64;
```

alignment_scores:
```
               Quality:  263.50
               Ratio:      0.620
               Percent Similarity:  41.222      Percent Identity:  19.011
```

alignment_block:
US-09-664-641-10:
US-09-664-641-10 x Q92547

Align seq 1/1 to: Q92547 from: 1 to: 1550

```
751 pValLysAspAlaLeuAlaLeuAlaLeuGluThrProGlyArgProSerGlnG 768
        ::::|||    :::||||
2312 AT................................TTAGTTTTAAATCTT 2328
                                        |||| |||||
768 InLysArgLysProSerThrProLeuSerGluValAlaIleValLysAsnLeu 784
    :::||||:::|    |||||
2329 TTACAGCTTGGAGAGTTCCCTTAAAAGTGTCTGCGAGAGTTTGATGAG 2378
                                |||  |:::::|:
785 ......GlnLeuAlaLeuAlaAsnSerSerArgAsnAlaValAlaI 797
      |||:|   :::
2379 TATAAGACTACCTCCCAAACTGAAACAGAATGAAGAAGTAGCTAATGCCAGC 2428
     |||:|:||||||||
797 aLeuLeuSerAlaSerProGlnLeuGluLysGlu............... 806
    ||:||:|||||||||
2429 CTTCTTCCCAAAGACCAGAATTGAAGACGTACCACTCCCACTAAAAAG 2478
        ::::::||:|||||    ||||    |||
807 ......AlaGlnSerGluLysGluValThrLeuProLeuLeuLysLys 820
       ||:|||  |||    |||
2479 CTTAACTCCAGAATTGACCCTTTGTTCCTTTCACTCGGATTCGAGCCTGT 2528
       ::::|:||     ::|||||    |||
821 ValAlaValCysValSer............................... 826
    ::|||  ::
2529 CCAGGTTCAACAGTATATTATTAAGAAGCTCTACATTCTTGGTGGAGGAGGTTG 2578
                                           :::||:
827 ......LysSerLeuSerLysLysGlnSerGlyLeuAsnGlyIleLeuAlaA 841
       |||||:||   ||    |||
2579 CGGGAGTCTGCA........................CAGAAGTGCACACAC 2604
      ||||||:|                           |:::
841 laSerLeuGlyAlaAspThrArgTrpSerProPheAspGluThrValThrHis 857
    :|||    ::|||||||    |||:|
2605 CTCATT...........GCCAGCAAAGTGACTGCGCACCGTGAAGTTCCT 2642
     ::||||           |||||:|||||
858 PheLeuTyrGlnGlyAlaTyrProAsnAspThrAsnArgGluThrTyrLysSerVa 874
    ::|    |||||    :::||||
2643 GACGGGCGGATTTCTGTGCTGGAAGACACATACTGCCAGAGTGGGCTGGAAG 2692
        :::|||    |||||
874 aLeuGluArgGlyVal..........HisIleValSerGluHisTrpLeuLeuA 889
    :::|||    ||:|||||
2693 AATGGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAACTAC 2733
    |||       |||:|||    |||
889 spCysAlaGlnCysLysHisLeuProGluSerLeuLeuTyr 902
```

Right column alignment:
```
234  AATGCGACTAGCCTCACACATAAATCTCAGAGGATGGGGACAAATCCAGAGGT 283
           ||:::|:    ::|  |||
168  AsnValSerValThrHisLeuIleAla.........GlyGluVa 179
     |||    |||
284  GGGAGGAAGCTCGGGAAGTC..............TTTGACTTACCTGTTGTTAA 321
     ||||:|  ::::                      ||||||
179  lGlySerLysLysTyrLeuGluValAlaLysAlaAlaAsnLeu..LysLysProIleLeu..eLeuL..196
             :||:|||
322  AGCCTTCTTGGGTGGATTCTG.............................TCCGTT 347
     ::::|||||||:|||:                                 |||||
196  euProSerTrpIleLysThrLeuLysThrLeuGlnGluLysLysIle 212
        |||:|||
348  CAGTGTGGGAACTCTTCTGCCCAGTAAAATGGTTTTCTCCAGAATCATGTCA 397
               |||:||    ||:|:|||    |||
213  ThrArgTyrThrAspIleAsnMetGluAsnPhe................LysCysSpr 226
     :::|||    :::|||
398  GATTTTTTTGGAATCACTGCCTGCTTTCTCAGGGTGTTGATACAGAGCT 447
     ::|||      |||:||::|||
226  oIlePheLeuGlyCysIleIleCysValThr.................. 236
        :::|||||
448  GGAGCTCTTTGTTGGAGTCTTCCAGAGCTCTCCCAAGGAGAGGTAGGGAA 497
       |||||:||    |||:|||||
236  ................................................. 236
498  GGGAGCTTGTCCAGGCGAAGTTGGGAAGCACACAGAGATCATCTGCCTTCTT 547
          |||||    :::|||
236  ................................................. 236
548  CTGACCCGGTATTGATTGCAGGCGTGAGGCCCTCGTTGTAAGTGCTGGGTG 597
                       :::||||    :::|||
237  ..............GlyLeuCys..GlyLeu................... 241
598  TCATCTGAAGCAGCAAGTGCCCTGTGGGGTTCTGGGTTACGGTTCTATGGGGGG 647
           |||||||||    :::|||
242  .............AspArgLysGluValGlnLeuGlnLeuThrHisValLysHisGlyVal 255
498  AGATTGC........CAGCTAACCCTCAATAAGAAAATGCACCGCATTTGATTG 691
             |||||||    ::::||||
255  yGlnTyrMetGlyValGlnLeuGluLysMetAsn..GluCysThrHisLeuIleV 271
692  TTCCAGAGCCAAAGGGGGGGAAAAATACGAATGTGCGTTTAAAGCGAGCAAGT 741
     ||        |||||||||||    :::|||
271  alGlnGluProLysGlyLnLysTyrGluCysAla..LysArgTrpAsn 286
742  ATTAAAAATGTGACTCCTGACTGGGGTTCTGGATTGGGATTCACAGAGAAAC 791
          ||||||||    |||:|||
287  ValHisCysValThrHrgLnTrpPhePheAspSerIleGluLysGlyPh 303
792  CAAAAAGGACGGAAGCATTTATCATCCTGCGTCGAATAATTATATGAAGAGG 841
     :::|||||||    |||:|||
303  eCysGlnAspGluSerIleTyr........................... 311
842  AAGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGAAAATGAGGAACAAGAT 891
     ||  |||||||||||||||||||    :::|||
311  ysThrGluProArgProGluAlaLysThrMetProAsnSerSerThrPro 327
892  TCTCAGAATGAGGGGTAGTACAGATGAGAGAGTACAAGCCCTGCCAGGTCTCA 941
     :::||:|:|||    |||:|||
328  ThrSerGlnIleAsnThrIleAspSerArgThr............... 338
942  AGAAGGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCCACACTG 991
```

```
339  .........................................      581  eSerGlnLysSerPheLeuValLeuGlyPheSerGlnAsnGlu....  596
                        ::|||  :                                     ||||||
          LeuSerAspValSerAsnIles 346            1766 AGCAACTGCTGGCCACCTGGAAAAGGAATAATCCAGCGACATGGGCACT 1815
992  AAAAATCTAAAGGGAGGAATTAAATGTTCAGAATTCATCACCG                      ::: ||||||||||:::|   ||
          :::||||                               597  ....SerAsnIleAlaSerIleIleIlelysGluAsnAlaGlyLys 609
346  erAsnIleAsnAlaSerCysValSerGluSerIleCysAsnSerLeuAsn 362

1042 GAAAAACAGGAGAGAAATTTAAACTGGACCCGCGCCAAGTCCCACAGTT 1091   1816 GTTGACCCCACCTTCACGAGTGCGATGCAACGCAC............. 1848
          :::|||| |||  ::::|| | :::::::||                              ::: |||
363  SerLysLeuGluProThrLeuGlu.                        610  Ile...Met.SerLeuLeuSerArgThrValAlaAspTyrAlaValValPr 625
          AsnLeuGluAsnIle 375

1092 AGCTGCCACCAAAAACCCACCGCCTCCCTCAGGGGAAAAGGACCCTGGCTTC..A 1138  1849 ...CTTCTCTGTGGAGAGTCAAGTCAGGCGGCGTATGGACAGCAGCAATAA 1894
          ::::|||                                               |||||| :::::||||
375  uAspValSerAlaPheGlnAlaProGluAspLeuLeuAspGlyCysArg 392   625  oLeuLeuGlyCysGlyValGlnGlnAlaThrValGly.............. 636

1139 TTAACTTGTGTGCCAATGTCCCCGTCCCAGGTTACATTTGCCCCT 1188   1895 GAGAAAGAAAGAGAATGTGTTACTGCACACTGGTTTAAACACAGTCTTAAAA 1944
          ::  ||||||                                           ||||||  |||
392  letyrLeuCysGly 396                                 637  GluValThrAsnThrTrpLeuValThrCysIleAsp 649

1189 GAGGTCCGGGGTAATTTAATGGCTCCTGGTGACAAAAAACCTCCAAAGTTCTCA 1238   1945 AAAAAAAAAATGGTA...CCGCGCGACACCGGCCCTTCACTTCCCAGTGGC 1991
          :::||||||||                                         ::: |||
397  PheSerGlyArgGlyLeuAspLys.. 404                    650  TyrGlnIleThrLeuPheAspProLysSerAsnProLeuPheThrProValPr 666

1239 AAGATACGAAATGAATAGCTTACCTGGAGTCGTACCGGACACTGAGGA 1288   1992 CTTCCCACCGAGGAGAAAGCCATGTTCACCACGCATATTATTTCTGTGACTG 2041
          :::||||| |||   |                                     |||  ::::||||
405  ..LeuArgLeuIleA 410                               666  oValMetThrGlyMetThrProLeuGlnAspCysValIleSerPheSerG 683

1289 ATATTACTAATAATGATAGCTTGACATTCAGCAGATGAAACCGGCCATCAAAATGTA 1338   2042 GATTTGTTGATAGTGACAGAGATGACCTAAAAATTAAGGCTTATTTGGCA 2091
          :::||||||||                                         |||  ::::||||
410  snSerGlyIyGlyIyGlyValArgPheAsnGlnLeuAsn.. 424       683  LnCysAlaGlyIyAlaGluIyLysGluSerLeuThrPheLeuAlaAsnLeuLeu 699

1339 GCACA'ATCTTCACGACTCTTCAGCCACTTACGAGAAAAATTTAGACACACCA 1388   2092 GGTGCCAAAATATACGGCGTTTATCTATGCGGC................... 2121
          :::||||| ::|| |||                                     ||||||
425  ThrHisValIle.....  ValGlyAspIlyrAspSpGluLeuLysGlnPh 439   700  GlyAlaSerValGlnGluIyrPheValArgLysSerAsnAlaLysGlySg 716

1389 GGTGAATCACAGC...CAGCAGGGACTACAAAATGCCAATGCACTGCGT 1435   2122 ......AGCAACACAGTCCTCATCCTGTAAAGAACCAAACTGGTTTAAAGT 2164
          :::||||  |||                                         |||  ::::||||
439  errpAsnlysSerAlaHisArgProHisValGlyIlyAlaLysTrpLeu. 455   716  yMetPheAlaSerThrValThrProLeuGluLeuLysGlyIyArgIyGlyGlySerLyST 733

1436 TTAGCCAAGTGAAAGTGACTCCAGAGACACACAATGCTACAGCACCACAG 1485   2165 ATGAAAAAGGCCAAAAGAGTGGAGGAATACCCCTGTGTCAACGCGCCCAGTGGGCTT 2214
          :::||||| :::||| |||                                 |||||| :::||  :::
456  .....LeuGluCysPheSerLysGlyHisGlyHisHisGluGlnAlaAspGluLeu 469   733  yrGluAlaAlaLysValSerIrpAsnLeuProAlaValThrIleAlaTrpLeu 749

1486 ...CAGGCCAGCAGCAGCCAGCAGCACCCGGTTTTACACCTTCA 1529   2215 ...........................CGGCAACACATTCTTCT 2228
          :::|||:: |||                                          |||
470  TyrIleHisIleAlaAsnTyrGlnProValGluSerHis...Gl 485    750  LeuGluThrAlaArgThrGlyIyLysArgGlnAlaAspGluSerHisPheLeuIl 766

1530 GCCCCAGCAGATAATCCAGCTCCAGCAGCACCGGCAGCACAGATCTCTC 1579   2229 GGGAAAC............................TT 2237
          :::||||:  :::|| |||                                   |||
485  nProGluSerLysAlaIleAlaLeuLeuIlyLysLysLysAsnSerSerPheSerL 502   766  eGluAsnSerThrLysGluThrIyLysArgSerLeuGluGluThrGlulyS 783

1580 AGCAAACTTACCCCAGCAGCAGCCGGCGAT.................. 1608   2235 ............................................... 2235
          :::|||  ::|| |||
502  ysLysAspSpPheAlaProSerGlulyHisGlyGlnAlaAspGluLeu 518   783  lyIleAsnLeuAsnSerAspThrArgLaGluHisProGlyThrArgLeuGln 799

1608 ............................................... 1608   2236 ...........................................TT 2237
          ::
519  LeuSerGlnTyrGluAsnGlySerSerThrValValGluAlaLysThrSe 535

1609 ...........CCATTTTCACACGACAGCAGCAGCACCAA....... 1635   2238 TGAGGCACTGAGGCAGATT.........CAGTATAGTGGCTACACGGCATTCA 2281
          ::: |||:::::||                                       ::::||||::|||  :::||
535  rGlualArgProPheAsnAspSerThrHisAlaGlyProLeuAsnAsps 552   800  ThrHisArgIysThrValThrProLeuAspMetAsnArgPheGlnSe 816

1636 ..........CAGCAGCCACCACACTGCCTCAGCAGCATCAG 1668   2282 GT............................................. 2282
                                                                 :::   |||
552  erThrHisIleSerLeuGlnGluAsnGlnSerValSerHisCys 568   816  rLysAlaPheArgAlaValSerGlnHisAlaArgGlnValAlaAlaAlaS 833

1669 CTTTTTTGCACATGACAATGCAGCACACTGCAG...ATTCCAGAAAAAGGCTTCTT 1715   2283 ............................................... 2283
          ::: |||||||:::::||                                   |||
569  Val.........ProAspValSerThrIleThrGluGlyIyLeuPh 581   833  erProAlaValGlyIyGlnProLeuGlnLysGluProSerLeuHisLeuAsp 849

1716 ATTGGGATGTGTTTGCAATTGCCGATTATCCAGAGACACAGTGTCTGATA 1765   2284 ...............CTGCAGGAAGACCCAHTTGCCCCTCACCCAGC 2311
          :    |||                                             ::|:  |||  :::::||:
                                                            850  ThrProSerLysPheLeuSerLysAspLysLysProSerPheAs 866

                                                            2283 ............................................... 2283

                                                            866  pValLysAspAlaLeuAlaLeuAlaLeuLeuGluThrProGlyArgProSerGlnG 883
```

alignment_scores:

Quality:	229.00	Length:	725
Ratio:	0.727	Gaps:	21
Percent Similarity:	43.448	Percent Identity:	18.207

seq_name: sp_fungi:Q9P5J5

seq_documentation_block:
```
ID   Q9P5J5       PRELIMINARY;   PRT;   838 AA.
AC   Q9P5J5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CONSERVED HYPOTHETICAL PROTEIN.
GN   B23L21.340.
OS   Neurospora crassa.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Sordariales; Sordariaceae; Neurospora.
OX   NCBI_TaxID=5141;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Schulte U., Algn V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA   Nyakatura G., Mewes H.W., Manhaupt G.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   German Neurospora genome project;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356172; CAB91702.2; -.
KW   Hypothetical protein.
SQ   SEQUENCE   838 AA;   93662 MW;   F7558140TD68823D7 CRC64;
```

alignment_block:
us-09-664-641-10 x Q9P5J5

Align seg 1/1   to: Q9P5J5   from: 1   to: 838

```
1156 GTCCCACCCGTCCCCAGGTAACCATTTTTGCCCCTCCGAGGTCCGGGGTAATTT 1205
     :||:||||||||:||:||:||||:||:::|||::    |||    |||
78   IleProValValAsnSerAsnIlePheLysIleThrLeuAlaArgAsnLy 94

1206 AATGGCTGCTGGACAAAACCTCCAAAGTTCTTGAAAGATCAGAAAATGATAG 1255
     ::|||    |||                    :::::||
94   sValAla.................................GlnValA 99

1256 CTACCTGCAGTCCCAGCCTGTTACGGACACTGAGGAAT...ATTACTAAT 1299
     ::||:||||    |||                    :::    |||
99   rgProTyrSerProAspProArgMetIlePheSerAsnVal...leLeuThr 115

1300 AATGCTGACATTCAGCAGATGAACCGGCCCATCAAAATGTAGCACATATCTT 1349
     ||||||||    :::||||||::||||||::||||:||||:    ::
116  CysAlaAspIlePheGluSerAspLysGluThrIleIleGlyAlaThrMe 132

1350 ACAGACTCTTTCAGCACCTACACCGGTTTTACACCTTCAGCCCCAGCAGTGGAATTC 1399
     :::||||||:|||||||||||:||:||:||||||    ::::||:|:|||
132  tAlaLeuGlyGlyMetGluSerLysAspLeuThrArgGlnThrThrHisI 149

1400 GCCAGCAGCAGGGACACATACAAATGCCAAATGCCACTGTGGTTAGCCAAGTGAA 1449
     :::||||||||||    ||:||||::|||||:||    :::
149  le................................................ 149

1450 GTGACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGCAGGCCACCA 1499
     ||||||||||||||||||||||||||||||||||||||
149  ..........CysAlaLeuSerMetAspHisGlyLysCysGlnGluVal 162

1500 GCAGCAGCAGCACCCGTGTTTTACACCTTCAGCCCCAGCAGTAATGCAGC 1549
     |||||||||||    |||    |||
162  aGlnLysLysAsnProLysCysLysIleValLeuProHisTrpPheAspA 179

1550 TCCAGCAGCAGCAGCACCAGCAGGATCTCTCAGCAACCTTACCCCAGCAG 1599
     :::||||||||||||||||||||||    ||||||
179  spCysPheArgLeuGlyArgArgIleSerGluAlaProTyrMetLeu.. 194

1600 CCGCCGCCATCCATTTCACAGCAGCAGCAGCAACAGCAGCCACCACC 1649
     :::||||||||||||||||||    :::
194  .............................................Pr 195

1650 ATCGCCTCAGCAGCCATCAGCTTTTTTTGACATCATCCAGCAGTGGAGATTC 1699
     :::||||||    |||    |||
195  oAsnProGlu................GlyProAspGluAspIleThrIleP 211

1700 CAGAAGAAGGCTTCTTATTGGGATGTGTGTTGCAATTT.....GCGGAT 1743
     :::||||||||:|||:||::||:|    ||||||:::    |||
211  roProSerGluAlaValGluGlyAlaAspThrSerValIleProSerAlaAla 227

1744 TATCCAGAGCAGCAGATGTCTGATTAAGCAACTGCTGGCCACCTGGAAAAGGAT 1793
     ::||||||||||||    :::||||||:||    :::
228  MetProGlyArgGluArgGluLysIleLeuValValPheAlaGlnLysLysVa 244

1794 AATC 1797
     :::
244  .... 244

1798 IMetIleSerGlnAspLeuProIleAsnAlaArgLeuArgAsnIleLeuA 261

1834 AGTCGGATGCACGCACCTTCTCTTGTGAGATGCAAGTCACGACGCCGTATGC 1883
     :::|||    :::||||||:||:||::||||:||||:||||||:||:|||
276  AspAlaCysAspMetTyrValCysGlnThrTyrArgAspGlyAspGluTyrIl 292

1884 ACAGGCAAATAAGACAAAAGAAAGAGATGTGTTACTGCCACACTGGTTAAACA 1933
     :::|||    |||                    |||
292  eArgAlaAlaGlnGlnGlnGlyLysAspValGlyAsnLeuAlaTrpLeuTyrT 309
```

```
1934 CAGTCTTAAAAAAAAAATGTA...CCGCCGCACCGAGCCCTTCAC 1980
     ::::::                      ||| :: |||  |||||||||
 309 yrLeuIleValHisAsnGluThrPhrArgProThrArgLeuHis 325

1981 TTCCCAGTGGCGCTTCCCACCAGGAAAAGCCATGTTCACAGCATATT 2028
     ::||||||::                ||| ::::|||  ::
 326 TyrProIle.....  ProArgAspGlyIleProGlyPheThrGlyMetIy 340

2029 ATTTCTGTCACTGGATTTGTTCATACTGACGAGACGATGACCTAAAATTAA 2077
     ||| :::||          ||| |||  |||||||||:::
 340 sIleThrLeuSerAsnTyrGlyIyAspAlaArgValTyrLeuGluAsnL 357

2078 TGGGCTTATTTGGCAGGTGCCAAATAATAACGGGTTATCTATGCCGCAGCAAC 2127
     |||||||||:||::          |||  |||||||||:::||||
 357 euIleThrAlaAlaGlyTyrAlaThrTyrThrLysThrMetLysAlaAspAsn 373

2128 ACAGTCCTCATCTGTAAAGACCAACTGGTTTAAGATGAAAAAGCCAA 2177
     |||||||:           ::: ||| ||| |||||
 374 ThrHisLeuIleThrAlaArgMetLeuAsnGlyLysCysGluAlaAlaIy 390

2178 ACAGTCGGAGGATACCCCTGTGTCAACCCGCAGTGC.........  2211
     ::: :::          ||  |||
 390 sGluTrpAsnIleGluIleLeuAsnHisLeuTrpIleGluAspSerTyrA 407

2211 ................  AGTCGCTACACG 2274
 407 laLysCysGluValGlnProPheSerAsnProArgTyrGlnHisPhePro 423

2212 ..........  CTTGGGGACATTCTTCTGAGGAAACTTT........ GA 2240
         ::::::||::|||::||        ::
 424 ProArgThrAsnLeuGlyValGluIleLeuGlySerThrPhePheAspGluGl 440

2241 GGCACTGAGCGCAGATTCAGTAT................  2262
     :                ||| ||| |||  ::
 440 nArgLeuArgGluValTyrPheProGlyGlyGluGluValLeuSerThrA 457

2263 .....................  AGTCGCTACACG 2274
 457 laAlaLysArgArgLysAlaAsnAspAlaAlaAsnLysAsnValTyr 473

2275 GCATTCAGTCTGCGAGGATCCAATTGCCCCTACCCAGCGATTTAGTTTTAAA 2324
     ||||||||:::|||         ::
 474 AlaSerAsnLeuAlaGluThrAspAlaGlyThrAlaAsnLysSerSerPr 490

2325 TCTTTTTAGATGCTTGGACAGTTCCCTTAAAAGTGTCTCGCAGAGTTGTTGA 2374
     :               ||||  |||||| :
 490 oLeuGlyLysThrAlaIleLysProProArgGlyAlaGlnProRsnLysG 507

2375 TGAGTATAAGACTACCTCCCCAAACTG.............  2406
     ::::||::|||::|              ||::||
 507 lnAsnIleGlnThrProAlaLysThrThrArgThrArgLysGlu 523

2407 AATGAAGTAGCTAAATGTCCAGCCTTCCCCCAAAGA.............  2442
     ||||::||||       |||:||||:::|||
 524 AsnAspThrProSerValMetSerSerArgSerAlaLysAspIy 540

2443 ................  GCCAGAATTGAAGACGTACCACCT........ C 2467
     ::                          |||||  :
 540 sAlaLeuSerLysLeuHisGlnLeuAlaProAspIleALeuTyrGluL 557

2468 CCACTAAAAANGCTAACTCCAGAATTGACCCCTTTGTTGCTTTCACTGGA 2517
     |||||||||  ||| |||         :
 557 ysGluLysLysArgThrAlaLysAspGlyProTrpProGlyLysArgAla 573

2518 TTCGAGCCTGTCCCAGTTCAACAGTATATATTTAAGAAGCCTACATTCTT. 2565
     ::: ||                   |||:||
 574 AlaAspGlnIleAsnAspArgAsnArgAlaLeuLysGluSerSerProAlaGl 590

2566 ....GGTGGAGAGGTTGCCGAGTCTGCACAGAAG.............  2595
         ||::||                   |||::|
 590 uAspGlyValGluGlnLysAspGluThrGlnLysLysArgProAlaLysLysA 607
```

```
2595 ................................  2595
                                     :::
 607 laArgValSerSerProGluMetArgIleCysLeuThrGlyTyrLys 623

2595 ................................  2595
 624 ArgTrpValAsnAspLysMetArgGluGluAlaAspArgIleValGlnAs 640

2596 ......... TGCACACACCTCATTGCCAGCAAAGTGACTCGCACCGTGA 2635
              |||  |||||||:::||||||||||:::|
 640 pAsnValProCysAspTyrLeuVaLAlaLeuProArgMetValArgThrMetL 657

2636 AGTTCCTGACGGCGCGATTTCTGTCGTCGAAGCACAIAGTGACGGCCAGAGTGG 2685
     |||||||||          :::||||||:::::||||||||||||||
 657 ysPheLeuValProCysLeuAlaArgGlyTyrProAspProcysIleIleSerSerAspTyr 673

2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGACGAGAACTACAT 2735
     ||| |||          ::: ||| |||:::
 674 ValThrAlaCysValGluIlaLaGlyValLeuProProLysAspTyrLe 690

2736 TCTCCGAGATGCTGAGGCAGAGAGTACTTTGTCTCTTCAGCTTGGAAGAAT 2785
     :|||  |||  |||||||||||:::|||:::|::||||
 690 uLeuValAspGlySerGluAspArgPheGlyValThrLeuGlnThrA 707

2786 CCTTAAAACGGGCACACGTTCTCCACTCTTAAGGCAAAAATTTTTAC 2835
     :::::::||:||||::|||||||::|||::|||:::|||
 707 laIleSerArgAlaArgAlaAsn......ArgGlyArgLeuGluTrp 720

2836 ATCACACCTGGAATCTGCCCA.............  2870
     ::::::||||              ::::::::::::
 721 GlyValProValPheCysThrGluIleLeuLysAsnGlyValGlnSerTy 737

2871 GAAGGCAATCGTAGACTGTGCAGGA 2895
     :::::::||::|::||||:::  |||
 737 rGlnThrIleAlaGluAlaAsnGly 745
```

Fragments

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 4, 2002, 11:09:22 ; Search time 6332.45 Seconds
                   (without alignments)
                   11830.650 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct........gaacttagaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
                   Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 1046326293 residues

Total number of hits satisfying chosen parameters: 3595312

Post-processing: Minimum Match 0%
                   Maximum Match 100%
                   Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : GenEmbl:*
        1: gb_ba:*
        2: gb_htg:*
        3: gb_in:*
        4: gb_om:*
        5: gb_ov:*
        6: gb_pat:*
        7: gb_ph:*
        8: gb_pl:*
        9: gb_pr:*
       10: gb_ro:*
       11: gb_sts:*
       12: gb_sy:*
       13: gb_un:*
       14: gb_vi:*
       15: gb_vrt:*
       16: em_ba:*
       17: em_pr:*
       18: em_fun:*
       19: em_hum:*
       20: em_in:*
       21: em_mu:*
       22: em_om:*
       23: em_or:*
       24: em_ov:*
       25: em_pl:*
       26: em_ro:*
       27: em_sts:*
       28: em_un:*
       29: em_vi:*
       30: em_vrt:*
       31: em_htg_inv:*
       32: em_htg_other:*
       33: em_htgo_inv:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	2664	74.4	2732	9	HSU80735	U80735 Homo sapien
2	1495	41.8	1518	9	BC008328	BC008328 Homo sapi
3	1483	41.4	3882	10	AF104261	AF104261 Mus muscu
4	874	24.4	3771	5	AF172855	AF172855 Xenopus l
5 c	636.6	17.8	149592	2	AC073077	AC073077 Homo sapi
6	636.2	17.8	159041	2	AC093726	AC093726 Homo sapi
7	479.2	13.4	149992	2	AC073077	AC073077 Homo sapi
8	361.4	10.1	438	11	G30194	G30194 human STS S
9 c	353.4	9.9	120887	2	AC094546	AC094546 Rattus no
10	348.6	9.7	178419	2	AC074040	AC074040 Mus muscu
11	334.4	9.3	177205	2	AC079217	AC079217 Mus muscu
12	308.2	8.6	633	9	U23862	U23862 Human clone
13	263.4	7.4	690	6	G15998	G15998 human STS C
14	204.6	5.7	504	6	AX186240	AX186240 Sequence
15	186.8	5.2	579	6	AX187445	AX187445 Sequence
16	186.8	5.2	1342	2	AX184577	AX184577 Sequence
17 c	174.4	4.9	177205	2	AC079217	AC079217 Homo sapi
18	174.4	4.9	178419	2	AC074040	AC074040 Mus muscu
19	142.2	4.0	2122	9	AK056609	AK056609 Homo sapi
20	136	3.8	267	11	G24669	G24669 human STS W
21	136	3.8	462	11	G60373	G60373 SHGC-131148
22	135	3.8	209	9	U92824	U92824 Homo sapien
23	130.6	3.6	123554	9	AB023049	AB023049 Homo sapien
24 c	130.6	3.6	151228	9	AL662800	AL662800 Human DNA
25	130.6	3.6	200000	2	AP000512	AP000512 Homo sapi
26	130.6	3.6	221059	2	AL662797	AL662797 Homo sapi
27	123.2	3.4	444	11	G33227	G33227 human STS S
28	110.6	3.1	397	6	BD010479	BD010479 Diagnosti
29	105.6	2.9	5120	6	AX067148	AX067148 Sequence
30	105.6	2.9	5120	9	AF152102	AF152102 Homo sapi
31	105.6	2.9	1159	9	AF126748	AF126748 Homo sapi
32	105.2	2.9	1442	9	AF087653	AF087653 Homo sapi
33 c	105.2	2.9	1472	9	AF126749	AF126749 Homo sapi
34 c	104.8	2.9	720	9	AB007820	AB007820 Homo sapi
35	104.6	2.9	107637	2	AC094679	AC094679 Homo sapi
36	104.6	2.9	177298	2	AC068751	AC068751 Rattus no
37	103	2.9	234	6	AR002264	AR002264 Homo sapi
38	103	2.9	234	6	AR053098	AR053098 Sequence
39	100.6	2.8	119526	2	AC074468	AC074468 Sequence
40 c	100.6	2.8	138685	2	AC017582	AC017582 Drosophil
41 c	100.6	2.8	166244	3	AC093502	AC093502 Drosophil
42 c	100.6	2.8	173851	2	AC096966	AC096966 Drosophil
43	100.6	2.8	185400	3	AC104703	AC104703 Rattus no
44 c	100.6	2.8	343319	3	AE003536	AE003536 Drosophil
45	100.2	2.8	477	6	AR165922	AR165922 Sequence

ALIGNMENTS

RESULT 1
HSU80735
LOCUS   HSU80735   2732 bp   mRNA   linear   PRI 18-DEC-1997
DEFINITION Homo sapiens CAGF28 mRNA, partial cds.
ACCESSION   U80735
VERSION   U80735.1   GI:2555045
KEYWORDS   .
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2732)
AUTHORS   Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
            Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE   cDNAs with long CAG trinucleotide repeats from human brain
JOURNAL   Hum. Genet. 100 (1), 114-122 (1997)
MEDLINE   97369492
PUBMED   9225980
REFERENCE   2 (bases 1 to 2732)
AUTHORS   Margolis,R.L., Abraham,M.R., Gatchell,S.B., Li,S.H., Kidwai,A.S.,
            Breschel,T.S., Stine,O.C., Callahan,C., McInnis,M.G. and Ross,C.A.
TITLE   Direct Submission
JOURNAL   Submitted (02-DEC-1996) Psychiatry, Johns Hopkins Univ. Sch. of

```
COMMENT     Med., 600 N. Wolfe Street, Meyer 2-181, Baltimore, MD 21205, USA
            Contains a polymorphic region having an allele frequency of 145;
            1.00 (allele bp). In 32 chromosomes tested, heterozygosity was 0%.
FEATURES             Location/Qualifiers
     source          1..2732
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="7q; distal to WI-8540"
                     /tissue_type="brain"
                     /dev_stage="fetus"
                     /note="single strand sequence"
     CDS             <1..2236
                     /note="contains glutamine repeats"
                     /codon_start=2
                     /product="CAGF28"
                     /protein_id="AAB91434.1"
                     /db_xref="GI:2565046"
                     /translation="INNEEQDSQNEGSTDEKSSPASSQSQSPSGDQGFSPKSNTEKSKG
                     ELMFDDSSDSSSPEKQFRNLNWTPAEVPQLAAAKRRLPQGKEPGLINLCANVPVPGNI
                     LPPEVRGNIMAAGQNLQSSPSREMIATWSPAVRTLRNTINNADIQQMMRPSNVAHILQ
                     TLSAPTKNLEQQVNHSQQGHTNNAVLFSQYKVPETHMLQQQQAQQQQQHPVLHL
                     QPQQIMQLLQQQQQQISQQPYPQQPHPFSQQQQQQAQQQQQQHPVLHL
                     EGFLLGCVFAIADYPEQMSDKQLLATWKRIIQAHGGTVDPDSRVDARTFSVRVKSAA
                     RIAQAIRERKCVTAHWLNTVLLKKKMVPHRALHFPVFAPPGCKRPCGSOHIISVTGPV
                     DSDRDLKLMAYLAGAKYTGYLCRSNIVLICKEPIGLAYEKAKIWRIPCVNAQWLGDI
                     LGNFEALRQIYSRYTAFSLQDFPAPTQIHLVLNILLDAWRVPLKVSAELLMSIRLPPK
                     LKQNEVANVQPSSQRARIEDVPPTKKLTPELTFEPVLTGTEEPAVQVYQQIIKKLYILGG
                     EVAESAQKTHLIASKVTRTLKFLAAISVKHIIVPEWLEECFRCQKFIDEQNYILRD
                     AEASVLFSFSLEESLKRAHVSPLFKAXYFITPGICPSLSTMKAIVECAGKVLSKQP
                     SPRKLMEHKQNSSLSHILLISCHNDLHLCREYFARGIDVHNAEFVLTGVLTQTLDYES
                     YKFN"
     repeat_region   602..616
                     /rpt_type=tandem
                     /rpt_unit=CAG
     repeat_region   620..637
                     /rpt_type=tandem
                     /rpt_unit=CAG
     repeat_region   680..700
                     /rpt_type=tandem
                     /rpt_unit=CAG
BASE COUNT      816 a    612 c    605 g    695 t     4 others
ORIGIN
```

```
Query Match             74.4%;  Score 2664;  DB 9;  Length 2732;
Best Local Similarity   99.4%;  Pred. No.0;  Mismatches 14;  Indels 2;  Gaps 2;
Matches 2692; Conservative   0; Mismatches 14;  Indels 2;  Gaps 2;

Qy    874 gaaaatgaggaacaagaattctcagaatgaagggtagtacagagtagagaagtcaagccctgcc 933
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GAAAATGAGGAACAAGAATTCTCAGAATGAAGGGTAGTACAGAGTAGAGAAGTCAAGCCCTGCC 61

Qy    934 aggctccaagagaaggggtcctccagtgaccagagagtttcaacctaacaactgaa 993
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 AGCTCTCAACAAGCGTCTCCTTCAGTGACCACACAGTTTTCAACCTAACAACTGAA 121

Qy    994 aaatctaaaagggaaattaaatgttaatgatttcttcagattcaatcaacggaaaaacaagag 1053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 AAATCTAAAGGGGAATTAAATGTTTGATGATTTCTTCAGATTCAATCACCGAAAAACAGGAG 181

Qy   1054 agaaattaactggaacccgccgaagtccacagttagctgcagcaaaacgcaggctg 1113
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 AGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAAACGCAGGGCTG 241

Qy   1114 cctcaaggagaagagagcctggttgattaacttgtgccaatgtcccaccegtcccagt 1173
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CCTCACGGAAACAGAGCCTCGGGTTCATCAACTTGTCGCAAATGTCCCACCCTCCCAGT 301

Qy   1174 aacatttgccccctgaggtccgggggtaattaaatgagcaaaaacctcaaagt 1233
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 AACATTTTGCCCCCTGAGGTCCGGGGGTAATTTAAATGGCTCGCTGGACAAAAACCTCAAAGT 361
```

```
Qy   1234 tctgaaagatcagaaattgatagctacgtacggacactgaggaatatt 1293
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 TCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTTACGGACACTGAGGAATATT 421

Qy   1294 actaattaaatgtcgacattcagcagaatgaaccggcaatccaaatgtagcacatattctaag 1353
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 ACTAATTAATCCTGACATTCAGCAGAATGAACCCGGCCATCAAAATGTAGCACATATCTTACAG 481

Qy   1354 actctttcagcaccttacgaaaaaattagaaacagcaggtgaacacacagccaggggacat 1413
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 ACTCTTTCAGCACCTTACGAAAAAAATTAGAACAGCAGGTGAATCACACAGCCAGGGGACAT 541

Qy   1414 acaaatgccaatgcaagttgtagctacttagtcaagtgactccaagaagacacatgcta 1473
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 ACAAATGCCAATGCACAGTTGCTGTTTAGCCAAGTGAAGTGACTCCAAGAAGACACATGCTA 601

Qy   1474 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagccaggtttacacctcagcc 1533
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTTACACCTTCAGCCC 661

Qy   1534 cagcagaatgagctccagcagcagcagcagcagcagcagcagcaaccttaccc 1593
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 CAGCAGAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAACCTTACCC 721

Qy   1594 cagcagcagcccgcgcatccaattcttcacagcagcagcagcagcagcagcacccatcg 1653
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    722 CAGCAGCCCGCGCCATCCAATTCCACATTTCACAGCAGCAGCAGCAGCAGCAGCACCATCG 781

Qy   1654 cctcaagcagcagtcttttggacatgatccagcagcagtcagagattcaagaagaaggctc 1713
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    782 CCTCACCACGCACGCTTTTTTGGACATGATCCAGCAGTCAAGTCAACAGGAGAATTCC 841

Qy   1714 ttattgggattgtgtgtttgcaattgcggaatatatccagcagatgtctgataaggcaactg 1773
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    842 TTAATTGGGATGTGTGTTTTGCAATTGCGGAATTATCCAGCAGCAGATGTCTGATAAGGCAACTG 901

Qy   1774 ctggccacctgaaaagacacatggcgcacttgttgacccaccttacg 1833
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    902 CTGGCCACCTGGAAAAGGACATAATCCAGGCACCAAATGGCGCACGACCTTACG 960

Qy   1834 agtcgaatgcacgcacctctctgtgagaagtcaagtcaagccggta-tgcacaggcaat 1892
          |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    961 AGTCGAATGCACCTGGAAAAGGACGTAATCCAGGCAGCAAATGTTGA-CCCACCTTCACG 1020

Qy   1893 aaggagaaagaagagatgtgtttactgcgaacactggtcaaacacaagtcttaaaaaataaaaa 1952
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAGAGAAAGAAGAGATGTGTTTACTGCAACACTGGTTAAACACAGTCTTAAAGAAGAAGAGAA 1080

Qy   1953 aatggtaccgccgaccgagcccttcacttcccagtggccttcccaccaggagagaagc 2012
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AATGGTACCGCCGCACCGAGCCCTTCACTCCCAGTGGCCTTCCCACCAGGAGGAGAAGC 1140

Qy   2013 atgttcacagcacattatttctgtactggattgttgatagtgacagggatgacctaaa 2072
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 ATGTTCACAGCATATTATTTCTGTGACTGGATTGTTGATAGTGACAGGGATGACCTAAA 1200

Qy   2073 attaatggcttattggcaggtgccaaaatatacgggttatctatgcccgcagcaacacagt 2132
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ATTAATGGCTTATTGGCAGGTGCCAAAATATACCGGGTTATCTATGCCGCAGCAACACAGT 1260

Qy   2133 cctcaatctgtaagagaaccaactggtttaaaagtatgaaaagccaaagagtggagggatacc 2192
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCTCATCTGTAAGAGAACCAACTGGTTTAAAAGTATGAAAAAGCCAAAGAGTGGAGGGATACC 1320

Qy   2193 ctgtcaacgcccagtggcttggcgacattcttctgggaaacttgaggcactgagcca 2252
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTGGGAAACTTTGAGGCACTGAGCCA 1380

Qy   2253 gattcagtatagtggctacacggcattcagttcgctcagaggatccattttgcccctacccagca 2312
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATTCAGTATAGTGGCTACACGGCATTCAGTTCTGCAGGATCCCATTTTGCCCCTACCCAGCA 1440

Qy   2313 tttagttttaaatctttagatgcttgggagagttccctaaaagtgtctgagagttgtt 2372
```

```
Db   1441  TTTAGTTTTAAATCTTTTAGATAGTCTTGGGACAGAGTTCCTTAAAGTGTCTGCAGAGTTGTT
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2373  gatgagtataagactacctcccaaactgaaacagaatgaaagtagctaatgtccagccttc  2432
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAAGTAGCTAATGTCCAGCCTTC  1560
Qy   2433  ttcccaaagagccaggaattgaagacgtaccaacctaaaagctaacttccagaatt  2492
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  TTCCCAAAGAGCCAGCAGAATTGAAGACGTACCACACCTACCAACTTCCAGAATT  1620
Qy   2493  gaccccttttgtgtttcacttggattcaacagttcaacagtatattaagaa  2552
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  GACCCCTTTTGTGCTTTCACTTGGATTCAGACCGTGTCCAGGTTCAACAGTATATTAAGAA  1680
Qy   2553  gctctacaattccttggttgagaggttgtccggaggtctgcacagaagtgcacacctcattgc  2612
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  GCTCTACAATTCCTTGGTGAGAGGTTGCGGAGTCTGCACAGAAGTGCACACCTCATTGC  1740
Qy   2613  cagcaaagtgactcgcacccgtgaagtctctgcgtgaagcacatagt  2672
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  CAGCAAAGTGACTCGCACCCGTCAAGTTCTCTGCGGCCCATTTCGTCGTCGACAGCATAGT  1800
Qy   2673  gacgccagagtgggctggaagaatgctctcaggtgtcagaaggttcattgatgaggcagaacta  2732
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  GACGCCAGAGTGGGCTGGAAGAATGCTTCAGGGTGTCAGAAGTTCATTGATGAGGCAGAACTA  1860
Qy   2733  catttccgagatgctgaggcaggagtactttcttctcctttcagcttgaagaatcccttaaa  2792
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  CATTTCCGAGATGCTGAGGCGGGAGTACTTTTTCTCTTTCAGCTTGGAAGAATCCTTAAA  1920
Qy   2793  acgggacacagtctctccacctcttaaggcaagaaatatattttacatcacaccctggaatctg  2852
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  ACGGGACACACGTTTCTCCCACTCTTTAAGCCAAAAATATTTTTACATCACACCTGGAATCTG  1980
Qy   2853  cccaagtcttccacatatgaaggcaatcgtagagtgtgcaggagggaaaggtgttatccaa  2912
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981  CCCAAGTCTTCCACTATGAAAGGCAATACGTAGAGTGTGCAGGAGGGAAAGGTGTTATCCAA  2040
Qy   2913  gcagccatcttccggaagtcatggagcacaagcagaactcgagtttgttcggaaataat  2972
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  GCAGCCATCTTTCCGGAAGTCATGGAGCACAAGCAGAACTCGAGTTTGTTCGGAAATAAT  2100
Qy   2973  tttaatalacctgtaaaatgaacttcatttatgccgaggagggcataga  3032
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101  TTTAATAATCCTGTGTCAAAATGACCTTCATTTATATNCCCGAGGGAGGGCATAGA  2160
Qy   3033  tgttcacaatgcagagttctgactggaagtgctcactcaaacgtggactatgaatc  3092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  TGTTCACAATGCAGAGTTCTGTTCTGACTGGAAGTGCTCACTCAAACGCTGGACTATGAATC  2220
Qy   3093  atataagttaactgtgaaaatgaacctcattatgccgaagaatatcttgccagaggcatag  3152
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221  ATATAAGTTAACTGACGATGGGCGTCCTAGGTGTGCACTCCGTGCACTCCGTGCGGGGGC  2280
Qy   3153  tggctgtctggctggcgagggagctcgctgctcctcacatgctcttgttctccagctg  3212
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  TGCGCTGCTCTGGCTGGCGAGAGCTCCTCGCTTCCTGTTCCAGCTG  3332
Qy   3213  cttcctgggggatcagagtgtgaaagcaacagaagtgaatataatacactgccatctt  3272
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  CTTCCTGTGGGGATCAGACTGTGGAAGCAACAGGAAGTGAAGATAATAATACACTGCATCTT  2400
Qy   3273  tttaagatgtgcaatttatttctgaggaaacataaataatgtgttctgatattatgacttt  3332
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401  TTTAAGATGTGCAATTTTATTCTGAGGAAACATAAAATATNGTTTGTTGTATTATGACTTT  2460
Qy   3333  aagagcccacattaggtttatgcaggtttaaatgtttcacaaaact  3392
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   2461  AAGAGCCCACACATTAGGTTTTTTATGATTCATTTTGCCCAGTTTTTAATGTTTTCACAAAACT  2520
Qy   3393  gttacgggacttccaactagaaataaaatgggtgtaaataaagaccttgctatcctcctaaatt  3452
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521  GTTACGGGGACTTCAACTACAGAAAATAAAAATGGGTGTAAAATAAAGACCTTGCTATCTCCTAAATT  2580
Qy   3453  atggatgttaaagatttgaaaatgtttgtacttgatctatgtttttcttatacctgt  3512
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  ATGGATGTTAAAGATTTGAAATGTTTGTACTTGATTTTTTTATTTTTCTTATACTGT  2640
Qy   3513  tttctttatattgatatcttgcccacatttaaataaaatgtacttttgaactttaagaaa  3572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641  TTTCTTTTATATTGANANTCTTGCCCACATTTTAAAATAAAATGTACTTTTGNACTTTAAAAA  2700
Qy   3573  aaaaaaaaa  3580
           |||||||||
Db   2701  AAAAAAAAA  2708
```

RESULT 2

BC008328

LOCUS   BC008328   1518 bp   mRNA   linear   PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3503689, mRNA, partial cds.
ACCESSION   BC008328
VERSION   BC008328.1 GI:14298977
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1518)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (25-MAY-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parveen Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy
            George Yang, Scott Zuyderduyn, Marco Marra.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 23 Row: c Column: 4.
FEATURES   Location/Qualifiers
     source   1..1518
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503689"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     CDS   <1..1177
                     /codon_start=2
                     /product="unknown (protein for IMAGE:3503689)"
                     /protein_id="AAH08328.1"
                     /db_xref="GI:14298998"
                     /translation="TVLKKKKMVPHRALHFPVAFPPGGKRPCSQHIISVTGFVDSFRD
                     DLKIMAYLAGAKYTGYLCRSNTVLLCKEEPIGLKYAKAEWRIPCVNAQWLGDILLGNF
                     EALRQIQYSRYTAFSLQDPFAPTQHLVHLNILLDAWRPLKVSAEILLMSIRIJPKLKQNE
                     VANVQPSSKRARIEDVPPPTKLTFPELLFPVLGTGEPVQVQQIIKKLIYILGGEVAES
                     AQKCTHLIASKVIRTVKFLIAISVVKHIVTPEWLEECFRCQKFIDEQNYIIRDAEAEV"

```
LPFSFSLEESLKRAHVSPLFKAKYFYITPGICPSLSTMKAIVEGAGQKVLSKQPSFRKL
MEHKQNSSLSEIILISCENDLHLCREYFARGIDVHNAEFVLIQGVLTQTLDYESYKFN"

Query Match      41.8%;  Score 1495;  DB 9;  Length 1518;
Best Local Similarity   99.7%;  Pred No. 0;
Matches 1498;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy   1932  cacagtctttaaaaaaaaaaaaaatgtaccgccgcaccgagcccttcacttccagtggc 1991
           ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CACAGTCTTTAAACAGAGGAGAAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCAGTGGC 60

Qy   1992  cttccacaccaggagaggaaaccatgttcacagcacatattattatttctgtggatttgtga 2051
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CTTCCACCACCAGGAGAGGAAAGCCATGTTCACAGCACATATTATTTCTGTGACTGATTTGTTGA 120

Qy   2052  tagtgacagagatgacctaaaattaaatgcttattgggagtgcaaatatacggtta 2111
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  TAGTGACAGAGATGACCTAAAATTAAATGCTTATTGGGAGTGCAAATATACGGGTTA 180

Qy   2112  tctatgccgcagcaacacagtcctcatctgtaaagagaaccaactggtttaaagtatgaaa 2171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TCTATCGCCGCAGCAACACAGTCCTCATCTGTAAAGAGAACCAACTGGTTTAAAGTATGAAAA 240

Qy   2172  agccaaagagtgagagaataccctgtgtcaacgcccagtggcttggcgacattcttcttggg 2231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  AGCCAAAGAGTGAGGAGAACCCCTGTGTCAACGCCCAGTGGCTTGGCGACATTCTTCTTGGG 300

Qy   2232  aaactttggagccactgagggcaggattcagtatatgtcgctacacggccattcagtctgcagga 2291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AAACTTTGGAGCGACTGAGGGCAGGAATTCAGTATAGTGCTACACGGCCATTCAGTCTGCAGGA 360

Qy   2292  tccatttcgccctaccccagcacattagtttcttaaatctctttagatgcttggggagttcctt 2351
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TCCATTTCGCCCTACCCCAGCACATTAGTTTCTTAAATCTTTTAGATGCTTGGGGAGTTCCCTT 420

Qy   2352  aaaagtgtctgcagagttgtgtagtataagaactacctcccaaactgaaacgaatga 2411
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGAACTACCTCCCAAACTGAAACAGAAATGA 480

Qy   2412  agtagctaatgtccagcgcttcttcccaaagagcagccagaattgaaagacgtaccacactccac 2471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  AGTAGCTAATGTCCAGCGCTTCTTCCCAAAGAGCCAGAATTGAAGACGTACCACACTCCAC 540

Qy   2472  taaaagctaactccagaattgacaccctttgtgtcttcactggattcgagcctgtcca 2531
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TAAAAGCTAACTCCAGAATTGACACCCTTTGTGTCTTCACTGGATTCGAGCCTGTCCA 600

Qy   2532  ggttcaacagtatattaagaaagctctacattccttggtgtggagagagtttagcggagttgcaca 2591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GGTTCAACAGTATATTAAGAAAGCTCTACATTCCTTGGTAGGAGGTTGCGGAGTCTGCACA 660

Qy   2592  gaagtggcaacacctcattggcaagaaagtgactggcaccgtgaagttcctgacggcgat 2651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GAAGTCCACACACCTCATTGGCAGCAAAGTGACTGGCACCGTGAAGTTCCTGACGGCGAT 720

Qy   2652  ttctgtcgtcgaacacaatgtgacgcagagtgggctggaagaatgcttcagggtgtcagaa 2711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  TTCTGTGCTGTCGAAGCAGTAGTGACGCAGAGTGGGCTGGAAGAATGCTTCAGGTGTCACA 780

Qy   2712  gttcattgatgacgaacctacattctccggagtagctgaggcgcagaagtactcttctcttt 2771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  GTTCATTGATGAGGACGAACTACATTCTCCGGAGATGCTGAGGCGAGAGTACTTTCTCTTT 840

Qy   2772  cagcttggaagaatccttaaaacgggcacacgtttcctccactcttaaggcagaaaatattt 2831
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CAGCTTGGAGAATCCTTAAAACGGGCACACAGTTTCTCCACTCTTTAAGGCAAAATATTT 900

Qy   2832  ttacatcacactggaatctgcccaaagtcttccactatgaaggcaatcgtagaagtgtgc 2891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TTACATCACACTGGAATCTGCCCAAGTCTTTCCACTATGAAGGCAATCGTAGAGTGTGC 960

Qy   2892  aggaggaaaggtgttatccaagcaagccaatcttccgggaaggctcaatggagacaagcaga 2951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  AGGAGGAAAGGTGTTATCCAAGCAAGCCAATCTTCCGGGAAGGCTCATGGAGACAAGCAGA 1020

Qy   2952  ctcgaagtttgtcggaaataattttaatatctgtgaaaatgactcaatttatgccagaga 3011
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  CTCGAAGTTTGTCGGAAATAATTTTAATATCCTGAAAATGACTCAATTTATGCCAGAGA 1080

Qy   3012  atatttgccaggggcatagatgttcacagtgcaagagttcgttctgactggaqtgctccac 3071
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  ATATTTTGCCAGAGCATAGATAGTTCACAATGCAAGAGTTCGTTCTGACTGGAGTGCTCAC 1140

Qy   3072  tcaaacgctggactatgaatccaatagaagttcactgatggctcttaggtctcggcgtcatg 3131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  TCAAACGCTGGACTATGAATCCAATAGAAGTTCACTGATGGCTCTTAGGGCTCTGCCGTGCATG 1200

Qy   3132  tcgactcctgcggtgcgggggctggctctggctggcgaggggagctgctggcttcctca 3191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  TCGACTCCTGCGGTGCGGGGGCTGGCTCTGGCTGGCGAGGGGAGCTGCTGGCTTCCTCA 1260

Qy   3192  catgtctcttgtttccagtcgcttcctcgggggatcagcagtgtgaaggcaggaagacagat 3251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  CATGTCTCTTGTTTCCAGCTCGTTCCTCGGGGATCAGCAGTGTGAAGCAGGAAGACACGAT 1320

Qy   3252  ataataaatatactgcatctttaagatgtgcaatttcattctgaggaaacatataaatta 3311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  ATAATAAATAATTATACTGCATCTTTTAAGATGTGCAATTTCATTCTGAGGAAACATAAAATTA 1380

Qy   3312  tgtttgtattatgacttaaggagcccacacattaggttttatgattcattgccagttt 3371
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  TGTTTGTAGTTATATACTTAAGAGCCCACATTAGGTTTTATGATTCATTGCCAGGTT 1440

Qy   3372  tttaaatgtttcaccaaaactgttacgggactccacatagaaataaaaagggttgtaaataa 3431
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  TTTAAAATGTTTCACAAAAACTGTTACGGGACTTCAACTAGAAATAAAAAATGCTGTAAATAA 1500

Qy   3432  aga 3434
           | |
Db   1501  AAA 1503
```

RESULT   3

```
LOCUS       AF104261               3882 bp    mRNA    linear   ROD 14-JUL-2000
DEFINITION  Mus musculus Pax transcription activation domain interacting
            protein PTIP mRNA, complete cds.
ACCESSION   AF104261
VERSION     AF104261.1  GI:4336733
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3882)
  AUTHORS   Lechner,M.S., Levitan,I. and Dressler,G.R.
  TITLE     PTIP, a novel BRCT domain-containing protein interacts with Pax2
            and is associated with active chromatin
  JOURNAL   Nucleic Acids Res. 28 (14), 2741-2751 (2000)
  MEDLINE   20368635
REFERENCE   2  (bases 1 to 3882)
  AUTHORS   Lechner,M.S. and Dressler,G.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1998) Pathology, Univ. of Michigan, 1150 W.
            medical Center Dr., Ann Arbor, MI 48109, USA
FEATURES             Location/Qualifiers
     source          1..3882
                     /organism="Mus musculus"
                     /strain="FVB"
                     /db_xref="taxon:10090"
     CDS             272..3442
                     /note="contains BRCT domains"
                     /codon_start=1
```

ORIGIN
BASE COUNT   439 a   322 c   339 g   418 t

/product="Pax transcription activation domain interacting
protein PIIP"
/protein_id="AAD17923.1"
/db_xref="GI:4336734"
/translation="MSEPAPEVPEELFREVKYAVGDIDPQVIQLLKAGKAKEVSVNA
LASHIISEDGDNPEVGEAREVPDLPVVKPSWVTLSVQGALLPVNGFSPESCQIFFGI
TACLSQVSSEDRSALWALVTFHGGSQLNLNKKCTHLIVEPKGEKYERAVKRTSIKI
VIPDWVLDCVSEKRRKDEAFVHIPRLIIYEEEEEEEEGDNEEQDSQNEGSTEKSSVAS
SAVASPAEQQCSPKPRAFVSKGEIMPDDSDSSPEKQERSINWAPAEAPPLNTAQRRL
PQGKPGLINLCANWPVPGDILPPDMRGNLMAPQONLQNSERSEILGTTWSPAVRTLR
NTINNADIOQIINRPSNVAHIIQSLSAPTKSLEOQVARGGOQGHPNASAVLFOQAKGAPE
THVLQOHHPPOPQVQOOPPALIHOVQQPOPPHONASAVI.FOOAKGAPE
HQIIOFSOQILOPPQOPLHPQQQLQRLQOOOMQNAANHSQASQALLQHQVLPQOPLQSL
QOPPPQQOQOHNLLQOQOQQOPOOLORLQOOOQMQNAANHSQASQALLQHQVLPQOPLQSL
QPPPQQOQOOLFCHDPAVEIPEESFLLGCVFAIADYPEOMSDKOLLATWKRIIOAHG
GTVDPPTPSRCTHLLCASQVSSMYIQAI.RERKHCVTAHWLNTVLRKKKLMPPHRALHF
PVAFPPPGCKPCSOHI1SVTGFVDNDRDDLKIMAYLACAKYTGYLCRSNTVLICKEPSG
LKYEXAKEWRIPCVNAQWLGDILLGNFEALRQVVQYSRYTAFNMEPDFVPTHLVLGLI
DANRTEPVKVTAELLMGVRLXPKLRPNEVANIQLSSKRARIEDLPPPKKLPPELUPLV
LFTGEPVVQVQVYIKKLYILGGEVAECTKKCCTHLLASKVTRTVKFLTAISVKHIVTP
DWLRECHKRQTFIDEONVILRDAFARVLFPSFSLEESLKRAHVSPPFKYFYITPGIC
PSLATMKAIVECAGCKVLAKQPSFKKLMEHKQNKSLSEIILISCNDLHICREYFARG
IDVHNARFVLFTGVILTQFLDVESYKPN"

BASE COUNT　　968 a　　1058 c　　1012 g　　840 t　　4 others

ORIGIN

Query Match　　41.4%;　　Score 1483;　　DB 10;　　Length 3882;
Best Local Similarity　　68.2%;　　Pred No. 0;
Matches 2553;　　Conservative　0;　　Mismatches　643;　　Indels 546;　　Gaps 14;

Qy　175　accttaggcacaggttattcaagcttctcaagctggaaaagcgaagaagaagttcctaca　234
　　　　|| || |||||||||||||||||||||||||||||| |||||| ||||| |||||||||||
Db　339　ACATCCACCCCAGGTTATTCAGCTCTTCCTTAAAGCTGGAAAGCCAAGAAGTCTCCTACA　398

Qy　235　atgcaactagcctcacacaatatctcaagaggatgggacaaatccagaggtgggaaaagtc　294
　　　　||||||||||||||||||| ||||||||||||||||||||||||||||||||||| |||
Db　399　ATGCCCTGGCTCCTCACACATCATTTCAGAGGATGGGGACAAACCCAGAGGTGGGAGAAGCC　458

Qy　295　gggaagtctttgacttactgtgtaaagcctcttgggtgtgattctccgttcaagtgtg　354
　　　　||| ||||||| ||| |||||||| || |||||||||||||||||||||| |||||||
Db　459　GTGAGGTCTTTGACTTGACCGTTGTAA...CCTTCCTGGGTGACTTCGTGTCGTGCAAATGTG　518

Qy　355　gaactctctgccagtaaatggttt:tccca:agaatcatgtcagatttttggaatca　414
　　　　| ||| ||||||||||||| |||| ||| |||||| || ||||||||| |||
Db　519　GCGCTCTCCTGCCAGTCAATCGTTCTCTGCAGACTCCTGCAGATTTTTTTGGACTCA　578

Qy　415　ctgcctgccttctcaggtgtgatacaagctggaagtctcttgtggaatctccagag　474
　　　　||||||||||||||||||||||||||||||||
Db　579　CTGCCTGCCTTCTCA--------------　594

Qy　535　calctgcctcttcgaccggtattgatgcaggcctgaagccctcgtaatgtgctgg　594
　　　　---G　595
Db　595　---

Qy　595　gtgtcatctgaagacaagaagtgccctgtgggcttggacttagttctatgggggagattgc　654
　　　　|||||||||||||||||||||||||||| ||||||| |||| |||
Db　596　GTGTCACTGGAAGACACAGAAGTGCCCGTGGGCCTTGGTTACGTTCCACGGGGCAGTTGC　655

Qy　655　cagctaaccctcaattaagagaaatgcacgcatttgattgttccagagccaaaggggggggaaa　714
　　　　|||||||||||||||||||||||| |||| ||| |||||| |||||||| ||||||
Db　656　CAACTAAACCTTAACACAGAAGTGCACACACTTGATTGTCCCAGACCCAAAGGGGGAGAAA　715

Qy　715　tacgaatgtgctttaaagcgagcagagtattaaaattgtgactcctgactgggtctggat　774
　　　　|||||||||||||||||||||||||| |||| |||||||||||||||||||
Db　716　TACGAGCGTGCTGTCTCAAAAGAACAGGAGTATTAAGATTGTCACCCGGACTGGGTCTGGAC　775

Qy　775　tgcgtatccagagaaaaccaaaagacgaagcacattttalcatcctcgtctgattatttat　834
　　　　|||| ||||| ||||| ||| |||| ||| ||| || ||||||| | ||||| |||
Db　776　TGTGTGTCTGAGAAGGAGGAGGAGGATGAAGACGTTTTACCATCCAGCCCTGATCATCTAT　835

Qy　835　gaaggaggaagaggaagtgcccctgtgggcttggacttagttctatgggggagattct　894
　　　　|||||| ||| || |||| |||| | ||
Db　836　GAGGAGGAGGAAGAGGAGGAG-------------GGAGACAACGAGGAGCAGGA?TCC　886

Qy　895　cagaatgagggtagtacagagtcaggccctgccagctcaaa...ggggtctcct　954
　　　　| |||| ||| |||||||||| ||| ||||||||||||
Db　887　CAGAACGAGGGCAGCACC---TGAGAGTCCAAGCTCTCCAAGCTGGGGGCCCCT　943

Qy　955　tcaggtgacaggcagtttcacctaaatccaacctgaaaaatctaaagggaattaatg　1014
　　　　||| |||| |||| |||||||| ||| |||| |||
Db　944　--GCTCAGACGCCTGTCTCCCAAGCCACCCCAGGGTCTGAGGTGTTCAAAGGGAGAGTTGAAG　1000

Qy　1015　tttgatgatcttcagatcacacggaaacccaaagttcctgaaagatcagaaatgata　1074
　　　　||| |||| ||| || ||| ||||| |||| ||||| ||||||||| ||
Db　1001　TTTGATGACTCTTCACTCATCCCTGACAAAACCTCCAGAATTCTGACGGTCAGAAATACTA　1060

Qy　1075　gccgaagtcccacagttagctgcagcaaacccaggcgtgcctcagggagaagaacctagg　1134
　　　　||| | |||||||||||| ||||| |||||||| ||||||||| || ||||
Db　1061　GCTGAAGCCGCCCACCACTCAAACACAGCCCACGCGCCTGCCCAGGGAAAAGCCCAGGCA　1120

Qy　1135　ttgataacttgttgccaatgtcccacccgttcccaggtaacattttgcccctgaggtc　1194
　　　　| || ||| ||| | || ||| || ||| ||| || |||| ||||
Db　1121　CTCATTAACCTGTGTCGCCAAATGTCCCACCTTGTTCCNGGGGACAATTCTGCCCTCCTTGA　1180

Qy　1195　cgggggtaattaatggctgctggaacaaaacctccaaagttcctgaaagatcagaata　1254
　　　　|||||| |||| |||| ||| || ||||||||| |||
Db　1181　CGGGGTAACTTGAATGGGCCCCTGACAAAAACCTCCAGAATCTGACGGTCAGAAATACTA　1240

Qy　1255　gctaactgggtccagctgtgtacggacactcgaggaagaatatattactaataatgctgacacttcag　1314
　　　　|||||||||| ||||||||||||||| |||||| | ||||| ||| ||
Db　1241　GGGACCTGGAGCCCAGGCGTGTCCCGACGTTGCACAGTTCGGAGACAACTACATTGATATCCAG　1300

Qy　1315　cagatgaaccggccatcaaatgtagcacatatttacagactcttcaagcacctacgaaa　1374
　　　　|||| ||||| || ||||| ||| | || ||||||||||||||||| |||||
Db　1301　CAGATCAACCGGCCATCCAATGTTGCACATATTACAGACTCTTCAGCACCCTACGAAAA　1360

Qy　1375　aatttagaacagcaggtgaatcacagacgagggactaacaaatgccaaatgcagtgctg　1434
　　　　||||| ||||| |||||||||| ||||||||| | ||||||||||||||||
Db　1361　AGCCTAGAACAGCAGGTTGGGGAAGTTGCGGCCGGGAAGCACACCCAATGCCAGTGCCTG　1420

Qy　1435　tttagccaagtgaagttgactccagagagacacagctacaagcagcacaggcccag　1494
　　　　||| |||||||||||||||| ||| |||||||||||||||
Db　1421　TTTTGGGGCCAGGCTAAAGGTGCTCCCCGACCACACGTGTTGCACACGAGTCTCCCCCAG　1480

Qy　1495　cagcagcagcagcagcacccggtgtttacaacctcagcccagcagaaatgcagttcag　1554
　　　　|||||||||||||||||| |||||||||||||||||| |||| || ||||
Db　1481　CAGCTCAGCAGCAGCAGCCACCAGCCCTGCACCTCCAGCCAGTCATGCAGCTGCAGCAG　1540

Qy　1555　cagcagcagcagcagca-----------------　1571
　　　　|||||||||||||||||
Db　1541　CAGCAGCAGCAGCAGCAACAACAGCAGCAGCCTACCCCCACCACCTCCATCA　1600

Qy　1572　------------------------　1572

Db　1601　CATCAGTTCCGCAGCAAGTCCATCAGCATCAGTTCTCTCAGCAGCTCCAGTTTCCA　1660

Qy　1572　----------------gatc-----　1575
　　　　　　　　　　　　　　||||
Db　1661　CAGCAACCATTACACCCTCAGCAGCCGCATCAATTGCAGCAGCTGCAGCCCATTT　1720

Qy　1576　tctaggcaacttaccccagcagccgccgatccattctcagcagcagcagcagcaa　1635
　　　　||||| || |||| || ||||||| || ||||||||||
Db　1721　CAGCAGCAGCATGCCCGTGCAGCAGCAGTTCATCAGTTGCAGCAGCAGCTGCAGCAC　1780

Qy　1636　cagcag-----------------　1641
　　　　||||||
Db　1781　CACCAGCTTAGCACAGCTCCAGCAACAGCAGCAGCAGCAACCTTCTCCAGCAGCAG　1840

Qy　1642　-----------------------　1641
　　　　
Db　1841　CAACAACACCAGCAGCTCCAAACGCTTGCAGCAGCACAGAGATGCAGAATCAAGCAGCA　1900

Qy   1642 -- 1641

Db   1901 CACTTGAGCCAAACGTCCCAGGCACTACAGCCACCAGGTTCTACCCGAGCAGCCCTGCAG 1960

Qy   1642 --------ccaacacacatcgcctcagcagcatcagcttctttggacatgatcca 1686

Db   1961 CTGTGCCCTGCAGCCCCCACCCTCAGCCACAGCAGCAGCAGCTTTTTGGACATGACCCA 2020

Qy   1687 gcagtgggagattcccagagaagaaggcgttccttattgggatgtgttgttgcaattggggattat 1746

Db   2021 GCAGTGGAGAATTCCAGAGAAGAAACAGTTCTTTGCTAGAGTGTGTGTTTTTGCAGAGTAT 2080

Qy   1747 ccagacagagtgtcttgataagcaacttgctggccacctggaaaaggacacat 1806

Db   2081 CCCAGCACAGATGTCCGGTAAGCACGTCCTGCCCACCTGGGAGATGTTTTTGCAATTGCAGAGTCA 2140

Qy   1807 gtcagcaagcgcgtatgcacaggccaataagagagagaaagaagagatgttactgcacactgg 1866

Db   2141 GGTGGCACTGTGGACACCCACTTCACAAGCCGATGTACACAACGTCCTGCCACCTGGG 2200

Qy   1867 ttaaacacagtcttaaaaaaaaaagtaccgcgcaccgcaccccttcacttccca 1926

Db   2201 GTCAGTAGTATGTTACACACAGGCACTACGGGAGAGAAAGCGGTGTTCACCACTGG 2260

Qy   1927 tgtgcaggagaggagaaaggtgttatccagacggcagccaatctctttccggaagctcatgggagcacaag 1986

Db   2261 CTCAACACAGTTCCTCAAGAGAAGAACAGACGTCGATGTGCCCTCAATCGAACCCT 2320

Qy   1987 gtcacctcccacaggaggaaagcacatgttcacagaacagacattttcttctgactgactgatt 2046

Db   2321 GTGGCGTTCCACCTCGGGGCAAACCCTGTTCACACCCTGGGAGACATTTTTCTGACTGGATTT 2380

Qy   2047 gttgatagtgacaagagatgaacctaaaataataatggcttatttggcaggtgcccaaaatacg 2106

Db   2381 GTTGACAATGACAGAGATGAACCTGAAGTTAAATGCATAACTACCTGCAGGTGCCAAGTACACA 2440

Qy   2107 ggttatctatgccgcagcaacacagtcctcatctgtaaagaacaactggtttaaagtat 2166

Db   2441 GGGTCACCTGTCTGCCCTCTAGCAACACACAGTCCTCATCTGTAAAGAACCAAGCCGTTGGGAGCTGT 2500

Qy   2167 gaaaaaagccaaaggtggaggatacccgtgctgtcaaccccagtggctggcccgaacattctt 2226

Db   2501 GAAAAAGGCCAAAGAGTGGAGGATCCCATGTGTGAATGCACACTGGCTGGGAGACATTCTG 2560

Qy   2227 ctgggaaacttgagggcagattcagtagagtcgctacaacggcattcagtctg 2286

Db   2561 CTGGGGAACTTTGAGGCCTGGGCAAGTGCAGTGCAGCCGGTACACCGCCTTCAACATG 2620

Qy   2287 caggatccattgccctaccccaagcatttagtttcttagatgcttggagagtt 2346

Db   2621 CCGGACCCCTTCGTTCCCACTCCACTCGCTCTTGGGTTCTTCAGATGCCATGGAGGACC 2680

Qy   2347 cccttaaagtgtctgcagagttgtgatgagtataaagactacctccccaaactgaaacag 2406

Db   2681 CCAGTGGAAGGTGACTGCAGAGCTGTTGATAGGGTTGTAAGGATTACNTCCTAAGCTGAAACCG 2740

Qy   2407 aatgaagtagtaagctaccagatgctcctcccaaagagacgacgtcaccacct 2466

Db   2741 AATGAGGTCCGTACATCCAGCTGTTCTTCCTAAAAAAGAGACCAGAATTGAAGACTTGCCACCT 2800

Qy   2467 cccactaaaaagtcaactcccagaattgacccctttgtgcttttcactggattcgagcct 2526

Db   2801 CCCACTAAGAAGCTGACTCGACGCCTTTACTGCGCTCTTTACTGTGCTCTGGGGTTTGAGCCTT 2860

Qy   2527 gtccagqttcaacagtatattaaaaaagtctcacattcctggattggagagtct 2586

Db   2861 GTTCAAGTTCAGCAGTACAGCTGACTACACACTTAAGNAGACTGTACANTTCTTTCGGGGAGTTCCTGACG 2920

Qy   2587 gcacagaagtgcacaacctcaattgccagacaaagtgactcgcaccgtgaagttcctgacg 2646

Db   2921 ACGAAGAAGTGTACCACCTCATTGCCAGCAAGGCTGCACGCACGTGAAAATTCTTGACG 2980

Qy   2647 gcgattctgtcgtgaagcacatagtgacgcgcagagtggctggaagaatgcttcaggtgt 2706

Db   2981 GCCATCACTGTCGTAGTAAAGCACACATTGTGACTCCGGACTCCGGACTGGGCTGGAGGAGTGCGTTCAAGCGC 3040

Qy   2707 cagagagttcattgatgagcagaagaactacattctccggaagatgctgaggagaagtacttttc 2766

Db   3041 CAGACANTTTATCGATGAGCAGAGNACTACATTCTTCGAGATGCTGAGGCGAGAAGTGTCTTC 3100

Qy   2767 tcttcagctcttggaagaatccttaaaaacgggcacacgtgttctccacctcttttaaggcaaaa 2826

Db   3101 TCTTTCAGTCTCGCAAGAGTCCTTAAAGGCAGCGACCCANTCTCTCCCCTNTCTTTAAGACAAAG 3160

Qy   2827 tattttttacaacacaccactggaatctgcccaagtcttccaactatgaagggacat 2886

Db   3161 TATTTCTATATCAACACCACTGGAAGATANATCCCCGATGCTCCCAACNTANTGAAGGGACAATTGTAGAG 3220

Qy   2887 tgtgcaggaggagggagaaaggtgttatccaagcagcacaatctctttccggaagctcatgggacaag 2946

Db   3221 TGTGCAGGGGGGAGAAAGGTGCTAGCCAAGCACAATCTGCCAAGCAAACCTTCCAACCTANTGAAGGGCAATTGTAA 3280

Qy   2947 cagaactcgagttgttccgggaaataattttaatatcctggtgaaaatgaccttcattatgc 3006

Db   3281 CAGAANTAAGACTCTTTGCAGAGATCNATTTTTAATTATCCTGGTGAAGAACCTGAAGAGACTTATTCT 3340

Qy   3007 cgagaatattttgccagagcagccatagatgttcacatgcagaagttcgttctgactggaagtg 3066

Db   3341 CGCGAGTACTTTTGCCACGGGACANTAGATGCTGTTCACANTGCAGAGTTGTCCCTCNACTGGAGTA 3400

Qy   3067 ctcactccaacgctggactatgaatcatatataagttaactgtgatatccagctgtcttttcctgggggtatcagaactgtg 3126

Db   3401 CTCACACANGACAGCTGGACTANTGANTCATATAAATTAANTGNAACTGAGTCCANTACANCAGATGGGCCTCT---TGCTGC 3456

Qy   3127 gcatgtcgactcctgcggttgcgggggctgctggctgtgctggctggaggaaggaagctgtctgcgcctc 3186

Db   3457 TGCTAGCCTNCAGCCCTGGAGCACCAAAGGGTAGCTGTGTGGCGCAGCANTGGACCAGCAGCAGCTCCCTC 3516

Qy   3187 cttcacatgctcttgtttcttcctggggatcagactgtgaactgtggaagcaggagaaga 3246

Db   3517 CCCCGCAGGGGCCTTGTTTTCTGCCCTTTCCAGCTCTGCCGTTAGGGAGCACGAGGACCTCG---ACCAGGAGCA 3572

Qy   3247 cagatataataataactgcaatcttctttttttaagatgtgcaatctattctgagaaaaacata 3306

Db   3573 CAGANTGTAATACTG---ACTGCAICGTGTTAAGGATGCATACTTAATTANT-AGGGAAACANTA 3629

Qy   3307 aattatgtttttgtattatgactttttaagagccacattaggttttatgatttcatttgcc 3366

Db   3630 AATTATGTTTTNTGTATTATATGCC---TAAGAGCCACATTAAGTTTNTTCNGATTTANTTNTGCC 3687

Qy   3367 aggtttttaaatgttttcacaacaaactgttacgggacttcaactagaaataaaaatggtgta 3426

Db   3688 AGGTTTTNTAAATGTTTTCACAACAAACTGTTGCAANTTTNCATTNAACANAATAAAAANTGANTGCT 3747

Qy   3427 aataaaagaccttgctatctctaaatatgagtgttaaagatttgaaatgttttttgtacttt 3486

Db   3748 AAATANAGACTTTTGC---TGTNTNAANTNATNGGCTGTTNAAGANATNGAAAATGNTTTTGTACTTTT 3805

Qy   3487 gattattttatttctgttttcttcttttatatcttgcccacatttta 3546

Db   3806 GATTANTNTNTTANTTCNTTACAC----TTTTCTNTTNTANTTCNTNGATNAACTNGCTCACA-GTTAA 3860

Qy   3547 ataaalgtacttttgaacttag 3568

Db   3861 AIAAAATGTACTTTGAACTTAG 3882

```
ORGANISM  Xenopus laevis
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
          Xenopodinae; Xenopus.
REFERENCE  1  (bases 1 to 3771)
AUTHORS    Shimizu,K., Bourillot,P.-Y., Nielsen,S.J., Zorn,A.M. and Gurdon,J.B.
TITLE      Swift Is a Novel BRCT Domain Coactivator of Smad2 in Transforming
           Growth Factor beta Signaling
JOURNAL    Mol. Cell. Biol. 21 (12), 3901-3912 (2001)
PUBMED     11359898
REFERENCE  2  (bases 1 to 3771)
AUTHORS    Shimizu,K. and Gurdon,J.B.
TITLE      Direct Submission
JOURNAL    Submitted (27-JUL-1999) Wellcome/CRC Institute, Tennis Court Road,
           Cambridge CB2 1QR, UK
FEATURES             Location/Qualifiers
     source          1..3771
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /tissue_type="embryo"
     gene            1..3771
                     /gene="K14"
     CDS             1..3771
                     /gene="K14"
                     /note="nucleoprotein"
                     /codon_start=1
                     /product="Swift"
                     /protein_id="AAK55123.1"
                     /db_xref="GI:14164561"
                     /translation="MAEDEVKVPEDMFKDVKYFPVGDLDPKVVVMLKAGKAKEVSYNA
                     LATHIIAHGDNPPVGFAREVFDLPVVKPSWVILSRCGVRLPENGFSPESGQIFGV
                     TACLSQVSPDDRNSLWALTFPYGGDCQLSLNKKCTHLIVPHPKGNKYEYAFQRGSIKI
                     VTPDWDLVSEKTKKDEALYHPRLIIVYEEEEEEEANNEEDEQSNGSDDKYSSK
                     SSPVTSRDGSPVHNREESSPKRSSSDKIKSELMFDDSSDSSPEKQERNLNWTPAEVPQM
                     STAKRQLHQGAPGSERPDMAFRSPAVRIRNITNSADVQQVNRPSNVAHILQTLSAS
                     TKSLEQRVNHPQQQHPNAVLFGQVKPLTSEAQHLIQQSHPHHQQQQQEQNHPIVQLQ
                     PQQLLQQQQQQILTQQVPRQHQFPQVNQQHPTQLQFPQQQLHPQQQILHRPQQQTIQ
                     HFQQQHALQQQLHQLQQHLQPKPQTLQQNMQQQNLQQPNLQQILQHQQTIQQTPSQQ
                     QALQPAIQQQQMLQPNIQQQQTLQQSNLQQQTLPIIQQQQGLQQQNLQQSLQQIQQQM
                     QHLTPQQKQQIQQQQQQQMLQKQQLQSQLQQQMQTHVLQQGTQIGTQALQQQQIQNQ
                     ALQPQQNQVQTHLALQQHLQSQTLQQQHHVQGQTVQQQQTHLQIQRTLLQQQHIQTLQ
                     IFHQIPAPNQQHQIPPPQMLQQQTILQLQQQHQLQPOMGSVQQQSSLQPQQMQQQHK
                     HNLQQVQHQLQQLQQRMQQRQMALPQQIANQQPPQHQLHGHDPPQDSVEIPEDPFLLGC
                     VFAIADYPEQQMPDKQILATWKRIIQTHGGTVDAALSRHCTHLLCESQVSSMYAQALKE
                     RKRCITFAHWLNAVLKLKKMVPPYRALHFPVAPPGGKPCSQHIISVIGFVDSDRDDLK
                     LMGYLAGAKXTGYLCRSNTVLCKEPSGLKYEKAKEWRIPCVNALWLCDILLGNLEAL
                     RQIQHSRYTVFNLQDPLAPSPHLVTDLLDAWRMPLKVSSEVLMSIRMPLKPKQNEPA
                     VQPKRPRIEDIPPPTKKLSPDQTPHVIVFGFPDPLQVQQYIKKLYILGGEVADTAQKCT
                     HIVANKVKTKVKFIJTAISVAKIIVTPEHLDESFKSQKPAEEQNYILRDAEAEVLFCFS
                     LEESLKKAHVNPLFKGKYFYIIPGIJCPSLSTMKAIVECAGGKILTKQPSFRKIMEHKQ
                     NKRLAEIIILSCENDLLHCREYFPAGSVDVHINAEFVLIGVLTQALDYESYKIYT"
BASE COUNT     1189 a    931 c    792 g    858 t    1 others
ORIGIN
```

```
Query Match        24.4%;  Score 874;  DB 5;  Length 3771;
Best Local Similarity   71.8%;  Pred. No. 1.6e-180;
Matches 1189;  Conservative   0;  Mismatches  451;  Indels  15;  Gaps   3;

Qy   1460 agacacacatgtcagcagcagcagcccagcagcagcagcagcaccggtt 1519
          ||||| | |||  ||| |||| || |||  |||| ||| |||| ||  |
Db   2126 AGCAACAAAATGCAGCCNCAGATTCAGCAACCCAGATGCAGTCAGCAACACT 2185

Qy   1520 tacacctcagcccagcagatatagctccagcagcagcagcagcagatctctc 1579
          || || ||| ||| || ||  ||  ||||| |||||| |||| |   |
Db   2186 CCTCACTCCAGCCACACAGCAGAATGCAACACATAAACACAATTGCAGCAGCTTCAGCATC 2245

Qy   1580 agcaacccttaccccagcagcgcgccgcatccatttcacagcaagcgcaacagc 1639
          || ||  ||| ||||| || | || |  ||  ||| || ||| || ||||
Db   2246 AACTGCAACAACTTCAACACGCAGCAGCGGATGCAGCAACGGCAGAAATTACCCAACAGA 2305

Qy   1640 agcc---accaccaatcgcctcagcagcaataagctcttttggacatgatccagcagtggaga 1696
          | |||  |||||||| ||| | |||   || |||| |  | |||   |  || ||
Db   2306 TTGCTCAATCAGCAAACCACCACACACGCGGCCGCCGATGATGAATCCTGACGCCCACGGTTCGTCTGTGGAGA 2365
```

```
Qy   1697 ttccagaagaggcttcttattgggatgtgtgttgcaattgcggattatccagagaga 1756
          |||||| |||| | ||  ||||||||| ||  |  ||  ||| | ||| ||||| |||
Db   2366 TTCCTGAAGAATTTCTTTTTTGGGATGTGTGTTTGGGAATGCTAGGCTGAATTANCCAGANCAAA 2425

Qy   1757 tgtctgtataagacaactgtctggcacctgggaaaaggataaatccaggcacatgggggcactg 1816
          || | | | ||||  ||| ||| ||| || ||||||  ||  | ||| || ||| ||| ||||
Db   2426 TGCCCGATAAAACAACTACTTGCAACCTGGAAAACGGATAATAACAGACACCCAGGAGGTACGG 2485

Qy   1817 ttgaccccaccttcacgagtcgatcgcacgccacctt cctctgtgagagtcaagtcagcagcg 1876
          || ||||| |||| || || | |  | || | ||||  || ||||| |  || ||||| |||
Db   2486 TGGACGCACCACTTTCAACCAGTGCACTACTTCTCTGTGTCGAAAGCCAAAGTCACTAGCA 2545

Qy   1877 cgtatgcacaggcaataagagaaagaagagaaagaagtgttactgcacactggttaaacacag 1936
          |||||| | |  ||| ||| |||| | ||| ||||| ||| | |||| | || |||||  |||
Db   2546 TGTAAGCTCAGGCATTGAAAGAGAGAGGAAGCCAATGGTAGTAICACTGCTCAGTTGGCGAATGCAG 2605

Qy   1937 tcttaaaaaaaaaaaaaatggttacccgccgcaccgagaccctttcacttcccacagtggccttcc 1996
          |  | | ||||||||||||||| || ||| | |  ||||  || |||||||| |||  | || || ||
Db   2606 TATTGAAAAANAGANGAAAAATGGTGCGCGCCTCTCATTTGGCCGAGCTCTTCATTTCCCTGTACCATTTC 2665

Qy   1997 caccagagaggaaagccatgttcacagcacatattattctgtgactggaatttgttgataagtg 2056
          ||||| | || | ||  ||| ||||| | ||  || || || ||| | |||  | | || | || |
Db   2666 CACCAGAGGAGAAAGCCCTCGTCTCACAGCATATCATTCAGIGACCGGGTTTGTTGGTTGACAGTG 2725

Qy   2057 acagagatgaactaaaaattaatggctcattggcaggtgccaaaatatacgggttatctat 2116
          |||| |||| ||| |||  ||  || |||  |||||| ||  |   | || |||| ||| ||
Db   2726 ACACAGAGATGAACTGAAAACTAATGAAAATGGGGTACTTAGCAGGGCCCAAGTACACTGGATACCTGT 2785

Qy   2117 gcgcgacgcaaacagtcctcatcttgtaaaggaacacaactggttaaaagtgtataaaagcca 2176
          || |  |||||| ||  | || |  || ||   ||| ||| ||  | ||||| | |  |||
Db   2786 GTCCGCAGCAAACACTGTTCTCATTTGTTGAAAGAGCCGAGCCGGGCTTGAAGTACGAGAAGCTA 2845

Qy   2177 aagagtggaggatacccctgtcaacgcccagtggcttggcgacactctctgggaaact 2236
          ||| ||| |||| ||  | ||  ||| | |||||| | |||| | || ||| || ||
Db   2846 AAGAGTGGAGGATACCGTGTGTAAAATGCCCTTGTGTAATTTCTTGGCATATCTTCGGCATTGGAAACC 2905

Qy   2237 ttgaggcactgaggccagattcagtatagtcgactcacacggcatcagtctgcagatccat 2296
          || ||| ||| | |||| |||| | | || || | |  || | ||  | ||  |||| |||
Db   2906 TTGAGGCAATACGTCAGTATTTCAGCACAGCCGAATACAGTATTTAATCTTCAGGACCCTC 2965

Qy   2297 ttgccccctaccccagcgaattagttttta---aatcttagatgcttggagagttccccttaa 2353
          |||| | | ||| | || |  || |  | |   |  ||  || | || |||| || || ||
Db   2966 TTGCACCATCACCTCATCTGGTTACAGATCATCTCCTAGATGCCCTGGGAGAATGCCATTGA 3025

Qy   2354 aagtgtctgcagagttgtgatgagtatagagactaactcccaaactgaaacagaaatgaag 2413
          | ||| | |  | ||||| | || | | | | | || ||||   || |||| |||| |||
Db   3026 AAGTGTCTCCTGGGAAGTTCTGAGTGAGTTAGTTCGGAATGCCCTCTAAAAGCCGAACGAAAAATCAGC 3085

Qy   2414 tagctaatgtccagcctcttcccaaagaaggaccagaaatgaagacgtacacactcccacta 2473
          || || |||| | || || |  | |||| | |||| | |||| |||| | || || || |
Db   3086 CTAAACGCCCAGGATTGAAGAAGATAAACGGGCTCCAACTA 3136

Qy   2474 aaaagctaactcccagaattgacccctttgtgctttcactgagttgagccctgtccagg 2533
          |||||| |||| || || || || || | || | |||| || | ||| ||| | || ||
Db   3137 AAAAGCTGTCACCTGATCAGACACCCTCATGTTATATTCACTGGATTGAATCAAGANCCACTGGCAGG 3196

Qy   2534 ttcaacagtatattaaagaagtcctacatcttgtggaggagtgtctgcacaga 2593
          |||||| ||| |  |||| || ||||  ||| | || ||| |  | |  ||| ||
Db   3197 TCCAGCAAATATATCAAAGAAGACTGTATAATTCTTGGCGGAGAAGTAGCCGGACAACTGA 3256

Qy   2594 agtgcacacacctcattggccagcaaagtgactcgcaccgtgaagttcctgacgcgcgatt 2653
          |  || | | || || ||| ||| ||  | || | |  || || ||||| | | || || |
Db   3257 AATGCACTCACTTGGTTGCCAACAAAGTGACCCGGCAACTGTAAAGTTCTTGACTGCAATTT 3316

Qy   2654 ctgtcgtgaagcacatagtgacgccagagatggctggaagaatgctcaggtgtcagaagt 2713
          |||| | || | ||| | || | ||| | || |  ||  || |||| ||| | ||| |||
Db   3317 CAGTGGCAAAGCACATTGTTACCCCGAGTGGCTAGATGAGAGCTTCAAAATCCCAAAAGT 3376

Qy   2714 tcattgatgagcgagaactacatctccgagatgctgaggcggaaggaagtactttctctttca 2773
          |  || |||| ||| || || |||||| ||| |||| | | |  || |||||  | |  | ||
Db   3377 TTGCTGAGGAACAGAATTACATATTTGCGGGATCGTGAAGCCGGAAGTGCTGTTCGCTTCA 3436
```

RESULT 5

```
Qy    2774 gcttggaagaatccttaaacqggcacacgtttctccacictttaagcgaaaatatttt 2833
           ||||| || |||| |   |||   | ||| || | | ||  ||||  |||| |||||||
Db    3437 CPTTAGAAGAATCCCATAAAAAAGCGCACGTGAAATCCGCTTCAAGGGAGGAAGTATTTTT 3496

Qy    2834 acatcacaacctggaatctgcccaagtctttccactgaagggcaatcgtagagtgtgcag 2893
           |||||||  ||||||| | | ||||||||| | |||| || | |||| ||  ||| ||| |
Db    3497 ACATCACCCCTGGAAATTGCCCAAGTCTCTCGACAATGAAAGCCATTGTAGAATGTGCTG 3556

Qy    2894 gaggaaagtgttatccaaggccatcttccggaagtcatggagcacaagagaact 2953
           | ||||| | |||  | || || |  | ||||| || |||| ||| |||  | ||
Db    3557 GTGGGAAAAATACTTAACGAAGCAGCCGTCGTCTTTTCGGAAAAATTATYGGAACATAAGCAGATA 3616

Qy    2954 cgagttgtcgaaataactttaatatcctgtgaaaatgaccticattatgccgagaat 3013
           | || |  |  ||| ||| ||||||| |||| ||||| |||| ||| | |||| || |
Db    3617 AACGATTAACTGACATCGATTTTAATATCTCCGAAAAATGATCTTCATCTGTGCACAGAAT 3676

Qy    3014 atttgccagaggcatagaqtgttcacaatgcagaagttcgttctgactggagtgctcactc 3073
           | ||| | | |  | | | ||| |||||||||| | ||| ||| |||| |||| ||| | |
Db    3677 ACTTTGCTGGCACGTGGCATGTGCACAATGCAGAGAATTTGTTTTGACGGAGTTCTCACCC 3736

Qy    3074 aaacgctggactatgaatcatataagttaactga 3108
           || | ||||||||||||||| ||| ||||||| |
Db    3737 AGGCACTGGACTATGAATCAATAATAAGTTACTTGA 3771
```

LOCUS   AC073077   149592 bp   DNA   linear   HTG 17-NOV-2000
DEFINITION Homo sapiens chromosome 7 clone RP11-269M19, WORKING DRAFT
            SEQUENCE, 10 unordered pieces.
ACCESSION   AC073077
VERSION   AC073077.5 GI:11192192
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Nov 17, 2000 this sequence version replaced gi:11054999.

            --------------- Genome Center ---------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            Center project name: H_NH0269M19
            -------- Project Information --------

            -------- Summary Statistics --------
            Sequencing vector: M13; 18%
            Sequencing vector: plasmid; 18%
            Chemistry: Dye-primer ET; 18% of reads
            Chemistry: Dye-terminator Big Dye; 64% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 142511 bases at least Q40
            Consensus quality: 144983 bases at least Q30
            Consensus quality: 146462 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 148692; sum-of-contigs
            Quality coverage: 5.83 in Q20 bases; agarose-fp
            Quality coverage: 5.97 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

FEATURES   Location/Qualifiers
     source   1..149592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-269M19"
     misc_feature   1..4209
                     /note="assembly_name:Contig10"
                     clone_end:T7
                     vector_side:left"
     misc_feature   4310..25549
                     /note="assembly_name:Contig11"
                     clone_end:T7
                     vector_side:left"
     misc_feature   25550..42417
                     /note="assembly_name:Contig12"
     misc_feature   42518..60534
                     /note="assembly_name:Contig13"
                     clone_end:SP6
                     vector_side:left"
     misc_feature   60635..97012
                     /note="assembly_name:Contig14"
     misc_feature   97113..139971
                     /note="assembly_name:Contig7"
     misc_feature   140072..141186
                     /note="assembly_name:Contig15"
     misc_feature   141287..143081
                     /note="assembly_name:Contig6"
     misc_feature   143182..145560
                     /note="assembly_name:Contig8"
     misc_feature   145661..149592
                     /note="assembly_name:Contig9"

     *   1   4209: contig of 4209 bp in length
     *   4210   4309: gap of unknown length
     *   4310   25449: contig of 21140 bp in length
     *   25450   25549: gap of unknown length
     *   25550   42417: contig of 16868 bp in length
     *   42418   42517: gap of unknown length
     *   42518   60534: contig of 18017 bp in length
     *   60535   60634: gap of unknown length
     *   60635   97012: contig of 36378 bp in length
     *   97013   97112: gap of unknown length
     *   97113   139971: contig of 42859 bp in length
     *   139972   140071: gap of unknown length
     *   140072   141186: contig of 1115 bp in length
     *   141187   141286: gap of unknown length
     *   141287   143081: contig of 1795 bp in length
     *   143082   143181: gap of unknown length
     *   143182   145560: contig of 2379 bp in length
     *   145561   145660: gap of unknown length
     *   145661   149592: contig of 3932 bp in length

BASE COUNT   40421 a 34230 c 33544 g 40490 t 907 others
ORIGIN

Query Match   17.8%; Score 636.6; DB 2; Length 149592;
Best Local Similarity   98.6%;
Matches 642; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

```
Qy    708 ggggaaatacgaatgtgtcttaaagccgagccaagtattaaaatgtgactcctggggt 767
          ||| |||||||||||| |||||||||||| |||||||||||| |||||||||||| ||
Db  57998 GCAGAAATACGAATGTCCTTTAAAGCCGACCAAGTATTAAAAATTGTGACTCCTGGGGT 58057

Qy    768 tctggattgcgtatcagagaaaaccaaaaagacgaagcattttatcatcctgtctgat 827
          |||||||| ||||| ||||||||||||||| | |||||||||||||||||||  | ||
Db  58058 TCTGGATTCCGTATTCAGAGAAAAACCAAAAAACGACGAAGCATTTTATCATCCTTCGTCGAT 58117

Qy    828 tatttatgaagaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaca 887
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  58118 TATTTATGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAACA 58177
```

```
Qy    888 agatlctcagqalgaqgqtagtacaagtgagaagtcaagcqcctqccaqctctcaagaqg 947
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58178 AGATTCTCAGAATGAGGCTAGTACAGATCAGAGAGTCAAGCGCCTGCCAGCTCTCAAGAGG 58237

Qy    948 gtctccttcaqqlqaccaqcaqtttcacctaaatcccacactqaaaaatctaaagqqqa 1007
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58238 GTCTCCTTCAGGTGACCAGCAGTTTCACCTAAATCCCACACTGAAAAATCTAAAGGGGA 58297

Qy   1008 attaatgtttgatqattcttcaggatcatccacqgqaaaacacaggagagagaaattaaactg 1067
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58298 ATTAATGTTTGATGATTCTTCAGGATCATCCACGGAAAACACAGGAGAGAGAAATTAAACTG 58357

Qy   1068 gaccccggccqaagtcccaccaqttagctcqcagcaaaacgcaggctgcctcaggqgaaaga 1127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58358 CACCCCGGCCGAAGTCCCACAGTTAGCTCGCAGCAAAACGCAGGCTGCCTCAGGGGAAAGA 58417

Qy   1128 qcctgggttgattaacttgtgtgccaatgtcccacccgtcccagqtaacattttgccccc 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58418 GCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCC 58477

Qy   1188 tgaggtccgggtgaatttaatgggtgctgctggacaaaacctccaaagttcctgaaagatcaga 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58478 TGAGGTCCGGGTGAATTTAATGGGTGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGA 58537

Qy   1248 aatgatagctacctggagtccagctacgqacactgaggaatattactaataatgctga 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  58538 AATGATAGCTACCTGGAGTCCAGCTACGGACACTGAGGAATATTACTAATAATGCTGA 58597

Qy   1308 cattcaagacagatgaaccggccatcaaatgtagcacatatcttacagactct 1358
          |||||||||||||||||||||||||||||||||   ||  ||
Db  58598 CATTCAGCAGATGAACCGGCCGTCAAATGTAGCACACATGTGTAGACTCACTCTTT 58648
```

```
RESULT  6
AC093726/c
LOCUS       AC093726              159041 bp    DNA     linear   PRI 08-FEB-2002
DEFINITION  Homo sapiens BAC clone RP11-5C23 from 7, complete sequence.
ACCESSION   AC093726
VERSION     AC093726.3  GI:18497181
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 159041)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 159041)
  AUTHORS   Tomlinson,C., Cotton,M. and Phillips,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-5C23
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 159041)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 159041)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 159041)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:15624904.
            -------------------- Genome Center
```

            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
            -------------------- Summary Statistics
            Center project name: H_NH0005C23
            Drafting Center: WIBR

            NOTICE: This sequence may not represent the entire insert of this
            clone. It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            (http://genome.wustl.edu/gsc)
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,H., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR: pBACe3.6

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,H., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An Improved
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7/, send
            mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

            NEIGHBORING SEQUENCE INFORMATION:
            The sequence of this clone was established as part of a mapping and
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Actual start of this clone is at base position 1 of RP11-5C23;
            actual end is at base position 23039 of RP11-279022.

            Data from AC074285 and AC073077 was used to finish this clone,
            AC093726. Polymorphisms have been identified between AC07428 and
            AC093726.

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-476H24, 6 bp overlap; the
            clone sequenced to the right is RP11-279022, 2000 bp overlap.
            Actual start of this clone is at base position 23039 of RP11-279022.

            The clone RP11-5C23 contains a 112 bp insertion not sequenced in
            the right neighboring clone, RP11-279022. This insertion sequence
            is being submitted as H_NH0005C23_F1.

            The sequence of AC013583 has been incorporated into AC093726.

```
FEATURES             Location/Qualifiers
     source          1..159041
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-5C23"
                     /clone_lib="RPCI-11"
     repeat_region   1..1735
                     /rpt_family="ERV1"
     repeat_region   1694..1731
                     /rpt_family="(GGA)n"
     repeat_region   1879..2117
                     /rpt_family="ERV1"
     repeat_region   2531..2719
                     /rpt_family="ERV1"
```

```
repeat_region    2748..3093     /rpt_family="ERVl"
repeat_region    3097..3405     /rpt_family="ERVl"
repeat_region    3380..3489     /rpt_family="ERVl"
repeat_region    3504..3814     /rpt_family="ERVl"
misc_feature     3994..4270     /rpt_family="Alu"
                 /note="similar to EST AA078489 (NID:g1837963)"
repeat_region    4427..4555     /rpt_family="Alu"
misc_feature     4683..5105     /rpt_family="Alu"
                 /note="similar to EST BF212893 (NID:g11106479)"
misc_feature     4683..5182
                 /note="similar to EST BE748183 (NID:g10162175)"
misc_feature     5048..5406
                 /note="similar to EST B1034826 (NID:g14441452)"
repeat_region    5206..5410     /rpt_family="MER1_type"
repeat_region    5449..5927     /rpt_family="MER1_type"
repeat_region    5944..5989     /rpt_family="Ll"
repeat_region    5944..5989     /rpt_family="(TAGA)n"
repeat_region    6053..6755     /rpt_family="Ll"
repeat_region    6752..6784     /rpt_family="Ll"
repeat_region    6756..7064     /rpt_family="(T)n"
repeat_region    7055..7607     /rpt_family="Alu"
repeat_region    7808..8174     /rpt_family="Ll"
repeat_region    8181..9662     /rpt_family="MaLR"
misc_feature     9470..9608     /rpt_family="MaLR"
                 /note="similar to EST AW855220 (NID:g7950913)"
misc_feature     9664..9915     /rpt_family="MaLR"
                 /note="similar to EST AW855220 (NID:g7950913)"
repeat_region    9815..10188    /rpt_family="MaLR"
repeat_region    10327..10565   /rpt_family="Ll"
repeat_region    10566..10650   /rpt_family="Ll"
repeat_region    10651..11124   /rpt_family="Ll"
repeat_region    11125..11667   /rpt_family="Ll"
repeat_region    11668..12906   /rpt_family="ERVK"
repeat_region    12921..12945   /rpt_family="Ll"
repeat_region    13343..13613   /rpt_family="AT_rich"
repeat_region    14320..14447   /rpt_family="MaLR"
repeat_region    14451..14488   /rpt_family="MaLR"
repeat_region    14573..14728   /rpt_family="MaLR"
repeat_region    14729..15068   /rpt_family="MaLR"
repeat_region    15069..15109   /rpt_family="MER1_type"
repeat_region    15467..15633   /rpt_family="Alu"
repeat_region    15611..15633   /rpt_family="MaLR"
repeat_region    16642..17000   /rpt_family="(A)n"
repeat_region    16642..17000   /rpt_family="MER1_type"

misc_feature     18037..18430
                 /note="similar to EST AW958088 (NID:g8147771)"
misc_feature     18039..18514
                 /note="similar to EST BF939756 (NID:g12356999)"
misc_feature     18040..18557
                 /note="similar to EST AW341490 (NID:g6838116) hd10u04.x1"
misc_feature     18040..18510
                 /note="similar to EST A1807116 (NID:g5393682) we85g11.x1"
misc_feature     18040..18330
                 /note="similar to EST A1284657 (NID:g3228890) qu10q01.x1"
misc_feature     18040..18510
                 /note="similar to EST A1671887 (NID:g4851618) wb41c03.x1"
misc_feature     18041..18344
                 /note="similar to EST H9448631 (NID:g9309666)"
misc_feature     18041..18510
                 /note="similar to EST BE890032 (NID:g10076656) uw58g12.y1"
misc_feature     18152..18585
```

```
Query Match    17.8%;     Score 636.6; DB 9; Length 159041;
Best Local Similarity 98.68%;     Pred. No. 2.5e-128;
Matches 642; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy    708 gggggaaatacqaatgtgctttaaagcqagcqaagtattaaaattgtgactcctgactgggt 767
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50682 GGAGAAATACGAATGTGCTTTAAAGCGAGCGAAGTATTAAAATTGTGACTCCTGACTGGGCT 50623

Qy    768 tctggattgcgtatcagagaaaaccacaaaaggagcqaagcgcattcatcatcctcgtctgat 827
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50622 TCTGGATTCGTATCAGAGAAAACCACAAAAGGAGCGAAGCGCATTTATCATCCTCGTCTGAT 50563

Qy    828 tattattgaagaggggaagaggaggaggaggaggaggaggaggaggagaagatgaggaagaca 887
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50562 TATTTATGGAAGAGGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTGAGGAACA 50503

Qy    888 agattctcagaatgagggtagtaccagatgagaagctcaagccctgccagctctcaaggagg 947
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50502 AGATTCTC*CAATGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTCAAGCAAGC 50443

Qy    948 gtctccttcaagtgaccagcagtttcacctaaatccaacacactgaaaaatctaaagggga 1007
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50442 GTCTCCTTCAGGTGACCAGCAGTTTCACCTAAATCCAACCACACTGAAAAATCTAAAGGGGA 50383

Qy   1008 attaatgtttgatgattctcagattccatcacccggaaaaaacagggaggagaadatttaaactg 1067
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50382 ATTAATGTTTGATGATTCTCAGATTCATCACCGGAAAAACAGGAGNAGAAATTAAACTG 50323

Qy   1068 gaccccgcgccgaagtcccacagttagctgcagcaaaaacqagctgcctcaggggaaagga 1127
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50322 GACCCCGCGCCGAAGTCCCACAGTTAGCTGCAGCAAAAACGCAGCTGCCTCAGGCGAAAGGA 50263

Qy   1128 gcctggggttgattaacttgtgtgccaatgtcccaccgtcccaggtaacattttgccccc 1187
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50262 GCCTGGGGTTGATTAACTTGTGTGCCAATGTCCCACCGTCCCAGGTAACATTTTGCCCC 50203

Qy   1188 tggagttccggggtatttaatggctgctggtgacaaaaacctccaaagttctgaaagatcaga 1247
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50202 TGAGGTCCGGGGTAATTTAATGGCTGCTGGTGACAAAAACCTCCAAAGTTCTGAAAGATCAGA 50143

Qy   1248 aatgaatagctacctgaagtcgacggacactgaggaaatactacetaataatgctga 1307
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  50142 AATGAATAGCTACCTGAGTCCGAGTCCGGACACTGAGGAAAT1AATTACTAATAATGCTGA 5008.

Qy   1308 cattcagcaagtgaaccgggccatcaaatgtagcacacatctcacagact.ct 1358
          |||||||||||||||||||||||||||| ||||||||||||||||  ||| |||
Db  50082 CATTCAGCAGATGAACCGGGCCCATCAAATGTAGCACACATGTAGTCACTCTTT 50032
```

```
RESULT    7
          AC073077/c
LOCUS      AC073077    149592 bp    DNA    linear    HTG 17-NOV-2000
DEFINITION Homo sapiens chromosome 7 clone RP11-269M19, WORKING DRAFT
           SEQUENCE, 10 unordered pieces.
ACCESSION  AC073077
```

VERSION   AC073077.5 GI:11192192
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE   The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 149592)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT   On Nov 17, 2000 this sequence version replaced gi:11054999.
            -------------- Genome Center --------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0269M19
            -------------- Summary Statistics --------------
            Sequencing vector: M13; 18%
            Sequencing vector: plasmid; 64%
            Chemistry: Dye-primer ET; 18% of reads
            Chemistry: Dye-terminator Big Dye; 64% of reads
            Assembly program: phrap; version 0.990319
            Consensus quality: 142511 bases at least Q40
            Consensus quality: 144983 bases at least Q30
            Consensus quality: 146462 bases at least Q20
            Insert size: 167000; agarose-fp
            Insert size: 148692; sum-of-contigs
            Quality coverage: 5.83 in Q20 bases; agarose-fp
            Quality coverage: 5.97 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 10 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *   1   4209: contig of 4209 bp in length
            *   4210   4309: gap of unknown length
            *   4310   25449: contig of 21140 bp in length
            *   25450   25549: gap of unknown length
            *   25550   42417: contig of 16868 bp in length
            *   42418   42517: gap of unknown length
            *   42518   60534: contig of 18017 bp in length
            *   60535   60634: gap of unknown length
            *   60635   97012: contig of 36378 bp in length
            *   97013   97112: gap of unknown length
            *   97113   139971: contig of 42859 bp in length
            *   139972   140071: gap of unknown length
            *   140072   141186: contig of 1115 bp in length
            *   141187   141286: gap of unknown length
            *   141287   143081: contig of 1795 bp in length
            *   143082   143181: gap of unknown length
            *   143182   144560: contig of 2379 bp in length
            *   144561   144560: gap of unknown length
            *   144561   145560: gap of unknown length
            *   145661   149592: contig of 3932 bp in length.
FEATURES   Location/Qualifiers
     source   1..149592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="RP11-269M19"
     misc_feature   1..4209
                     /note="assembly_name:Contig10

                     clone_end:T7
                     vector_side:left"
     misc_feature   4310..25449
                     /note="assembly_name:Contig11"
     misc_feature   25550..42417
                     /note="assembly_name:Contig12"
     misc_feature   42518..60534
                     /note="assembly_name:Contig13
                     clone_end:SP6
                     vector_side:left"
     misc_feature   60635..97012
                     /note="assembly_name:Contig14"
     misc_feature   97113..139971
                     /note="assembly_name:Contig15"
     misc_feature   140072..141186
                     /note="assembly_name:Contig6"
     misc_feature   141287..143081
                     /note="assembly_name:Contig7"
     misc_feature   143182..144560
                     /note="assembly_name:Contig8"
     misc_feature   144561..149592
                     /note="assembly_name:Contig9"
BASE COUNT   40421 a   34230 c   33544 g   40490 t   907 others
ORIGIN

RESULT   8
G30194/c
LOCUS   G30194   438 bp   DNA   linear   STS 05-OCT-1996
DEFINITION human STS SHGC-36464, sequence tagged site.
ACCESSION   G30194
VERSION   G30194.1 GI:1593745

Query Match   13.4%; Score 479.2; DB 2; Length 149592;
Best Local Similarity 98.4%; Pred. No. 6.9e-94;
Matches 484; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy   3079 ctggacttagtgaataataataaagttaactgatggcgtctaggctgccgtgcatgtcgactc 3138
            |||||||| | | || |||||||||||||||||||||||||||||||||||||||||||||
Db   111411 CTCAACAAGCCTACATATATAAGTTTAACTGATGGCGTCTAGGCTGCCGTGCAATGTCGACTC 111352

Qy   3139 ctgcggtccgggcgttggctgctgtggcgagagaagctgctgcgcttcacatggtgctc 3198
            |||||| | || |||||||||||||||||||||||||||||||||||||||||||||||
Db   111351 CTCCGGTGTCGGGGCTGGCTGTCTGCCGTGGCGAGAGAAGCTGCTGCGCTTCACATGGTGCTC 111292

Qy   3259 atatactgcatctttttaagatgtgcaatttttattctgaggaaacatataaatatgttttg 3318
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   111291 TTCGTTTTCCAGCTGCTTTCCTCGGGGGATCAGACTGTGAAGCAGAGGAAGACAGATAATAA 111232

Qy   3319 tattatatgacttaagagcccacattaggtttatgatcacttgtgcaggttttlaaat 3378
            |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   111231 ATATACTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATATGTTTTG 111172

Qy   3379 gttttcacacaaactgttacgggacttcaactaggaataaaatggttgtaataaaagacctt 3438
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   111171 TAPTATATGACTTTAAGAGCCCACAATTAGGTTTATGATTCACTTTGCCAGGTTTTTAAAT 111112

Qy   3439 gctatctctaaattatggatgttaaagatttgaaagtgtttttgtacttgattattttat 3498
            ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   111051 GCTATCTCTAAATTATGGATGTTAAAGATTTGAAAGTGTTTTTGTACTTTGATTATTTTAT 110992

Qy   3499 tccttatactcctgttttccttttatatattgatatcttgcccacattttaaataaatgtactt 3558
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   110991 TTCCT-ATACTCTGTTTTCTTTTATATGATATCTTGCCCACATTTTAAATAAATGTACTT 110932

Qy   3559 ttgaacttagaa 3570
            ||||||||||||
Db   110931 TTGAACTTAGAA 110920

KEYWORDS   STS; STS sequence; primer; sequence tagged site.
SOURCE   human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS   Myers,R.M.
JOURNAL   Unpublished (1996)
COMMENT

           Contact: Richard M. Myers
           Stanford Human Genome Center (SHGC)
           Stanford University School of Medicine
           Department of Genetics, M-344, Stanford, CA 94305, USA
           Tel: 4157259687
           Fax: 4157259689
           Email: myers@shgc.stanford.edu

           Primer A: TATTTAAAATGTGGGCAAGAYATCA
           Primer B: TCGTGTAAAAYAAAGAACCTTGCTATC
           STS size: 130
           PCR Profile:

           Initial incubation: 94 degrees C for 90 seconds

           Denaturation: 94 degrees C for 15 seconds
           Annealing: 62 degrees C for 23 seconds
           Polymerization: 72 degrees C for 30 seconds
           PCR Cycles: 30
           Thermal Cycler: Perkin Elmer 9600

           Protocol:
           Template: 25 ng
           Primer: each 1 uM
           dNTPs: each 200 uM
           Taq Polymerase: 0.05 units/ul
           Total Vol: 10 ul

           Buffer:
           MgCl2:   2.5 mM
           KCl: 50 mM
           Tris-HCl: 20 mM
           pH: 8.3

           Prepared with primer pairs provided by Sandoz, derived from T88893
           -- Washington University/Merck EST sequence.

FEATURES   Location/Qualifiers
     source   1..438
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /map="7"
     STS   1..438
     primer_bind   20..149
     primer_bind   complement(125..149)
BASE COUNT   155 a   82 c   77 g   117 t   7 others
ORIGIN

Query Match   10.1%;   Score 361.4;   DB 11;   Length 438;
Best Local Similarity 94.4%;   Pred. No. 1.7e-68;
Matches 405;   Conservative 0;   Mismatches 20;   Indels 4;   Gaps 3;

Qy   3138 cctgccggtgcgggggtctggctgtcgcgaggagtcgtgtcgcctt-cctcacatgc 3196
           ||||||||||| |||||| ||||||| | |||||| |||| |||||| | ||||||||
Db   435 CCCCGCCGTGCGCGGGGCCTGCCGGCGNCGCCGGAGGAGCCCNCGCCGCCTTCCCTTCACATGG 376

Qy   3197 tcttgtttt-ccagctgcttcctggggat--cagaactgtgaacgccagaagacagatat 3253
           || |||||| | |||||||||| |||| |||||| ||| ||||||||||||||||||||
Db   375 CCCCGTTTTCCCAGCTGCTTNCCCTGGGGGATCCAGACTGTGCAACGACCAGGAAGACAGATAT 316

Qy   3254 aataaaatactgcatctttttagatgtgcaatttattctgaggaaacataaattatg 3313
           ||||| ||||||||||||||||||||| |||| |||||||||||||||||||||||||
Db   315 AATAAAATACTGCATCTTTTTAAGATGTGCAANTTTATTCTGAGGAAACATAAAATATG 256

Qy   3314 ttttgtattatatgacttaagagcccacattaggtttttatgatcattgccaggtttt 3373
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
REFERENCE   2  (bases 1 to 120887)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Dec 20, 2001 this sequence version replaced gi:15624381.
  JOURNAL   Unpublished
COMMENT     Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ----------------------------------------------------------------------
                Project Information
                Center project name: GAWE
                Center clone name: CH230-412I
            ----------------------------------------------------------------------
                Summary Statistics
                Assembly program: Phrap; version 0.990329first call to
            findPhrapList
                Consensus quality: 81355 bases at least Q40
                Consensus quality: 89305 bases at least Q30
                Consensus quality: 95167 bases at least Q20
                Estimated insert size: 51481; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.6x in Q20 bases; sum-of-contigs estimation
            * NOTE: Estimated insert size may differ from sequence length
            * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 74 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *      1  2144: contig of 2144 bp in length
            *   2145  2244: gap of unknown length
            *   2245  5011: contig of 2767 bp in length
            *   5012  5111: gap of unknown length
            *   5112  7135: contig of 2024 bp in length
            *   7136  7235: gap of unknown length
            *   7236  9223: contig of 1988 bp in length
            *   9224  9323: gap of unknown length
            *   9324 11509: contig of 2186 bp in length
            *  11510 11609: gap of unknown length
            *  11610 13474: contig of 1865 bp in length
            *  13475 13574: gap of unknown length
            *  13575 15562: contig of 1988 bp in length
            *  15563 15662: gap of unknown length
            *  15663 17829: contig of 2167 bp in length
            *  17830 17929: gap of unknown length
            *  17930 20836: contig of 2050 bp in length
            *  20837 20936: gap of unknown length
            *  20937 22986: contig of 2050 bp in length
            *  22987 23086: gap of unknown length
            *  23087 25197: contig of 2111 bp in length
            *  25198 25297: gap of unknown length
            *  25298 26932: contig of 1635 bp in length
            *  26933 27032: gap of unknown length
            *  27033 29184: contig of 2152 bp in length
            *  29185 29284: gap of unknown length
            *  29285 31122: contig of 1838 bp in length
            *  31123 31222: gap of unknown length
            *  31223 33089: contig of 1867 bp in length
            *  33090 33189: gap of unknown length
            *  33190 34621: contig of 1432 bp in length
            *  34622 34721: gap of unknown length
            *  34722 36131: contig of 1410 bp in length
            *  36132 36231: gap of unknown length
            *  36232 37852: contig of 1621 bp in length
            *  37853 37952: gap of unknown length
            *  37953 39448: contig of 1496 bp in length
            *  39449 39548: gap of unknown length
            *  39549 41342: contig of 1794 bp in length
            *  41343 41442: gap of unknown length
            *  41443 43126: contig of 1684 bp in length
            *  43127 43226: gap of unknown length
            *  43227 44363: contig of 1137 bp in length
            *  44364 44463: gap of unknown length
            *  44464 45992: contig of 1529 bp in length
            *  45993 46092: gap of unknown length
            *  46093 47391: contig of 1299 bp in length
            *  47392 47491: gap of unknown length
            *  47492 49188: contig of 1697 bp in length
            *  49189 49288: gap of unknown length
            *  49289 50333: contig of 1045 bp in length
            *  50334 50433: gap of unknown length
            *  50434 52192: contig of 1759 bp in length
            *  52193 52292: gap of unknown length
            *  52293 54222: contig of 1930 bp in length
            *  54223 54322: gap of unknown length
            *  54323 55865: contig of 1543 bp in length
            *  55866 55965: gap of unknown length
            *  55966 56998: contig of 1033 bp in length
            *  56999 57098: gap of unknown length
            *  57099 58889: contig of 1791 bp in length
            *  58890 58989: gap of unknown length
            *  58990 61298: contig of 2309 bp in length
            *  61299 61398: gap of unknown length
            *  61399 63560: contig of 2162 bp in length
            *  63561 63660: gap of unknown length
            *  63661 65400: contig of 1740 bp in length
            *  65401 65500: gap of unknown length
            *  65501 66539: contig of 1039 bp in length
            *  66540 66639: gap of unknown length
            *  66640 67683: contig of 1044 bp in length
            *  67684 67783: gap of unknown length
            *  67784 68898: contig of 1115 bp in length
            *  68899 68998: gap of unknown length
            *  68999 70929: contig of 1931 bp in length
            *  70930 71029: gap of unknown length
            *  71030 72545: contig of 1516 bp in length
            *  72546 72645: gap of unknown length
            *  72646 73833: contig of 1188 bp in length
            *  73834 73933: gap of unknown length
            *  73934 75016: contig of 1083 bp in length
            *  75017 75116: gap of unknown length
            *  75117 76334: contig of 1218 bp in length
            *  76335 76434: gap of unknown length
            *  76435 78199: contig of 1765 bp in length
            *  78200 78299: gap of unknown length
            *  78300 80037: contig of 1738 bp in length
            *  80038 80137: gap of unknown length
            *  80138 81196: contig of 1059 bp in length
            *  81197 81296: gap of unknown length
            *  81297 82762: contig of 1466 bp in length
            *  82763 82862: gap of unknown length
            *  82863 84079: contig of 1217 bp in length
            *  84080 84179: gap of unknown length
            *  84180 85257: contig of 1078 bp in length
            *  85258 85357: gap of unknown length
            *  85358 86571: contig of 1214 bp in length
            *  86572 86671: gap of unknown length
            *  86672 87945: contig of 1274 bp in length
            *  87946 88045: gap of unknown length
            *  88046 89172: contig of 1127 bp in length
            *  89173 89272: gap of unknown length
            *  89273 90314: contig of 1042 bp in length
            *  90315 90414: gap of unknown length
            *  90415 91836: contig of 1422 bp in length
            *  91837 91936: gap of unknown length
            *  91937 93103: contig of 1167 bp in length
            *  93104 93203: gap of unknown length
            *  93204 94356: contig of 1153 bp in length
            *  94357 94456: gap of unknown length
```

RESULT 10

```
Query Match
*  9457 95931: contig of 1457 bp in length
* 95914 96013: gap of unknown length
Host Local Similarity    9.9%;   Score 353.4;  DB 2;
Matches 492; Conservative  74.0%;  Pred. No. 2.2e-66;
Matches 492; Conservative  0; Mismatches 161; Indels 12; Gaps  3;

Qy   682 cattlgalttgttccagagcaggcaaagggggggaaatacgaatgatgtgctttaaagcgagcaagt  741
         |||| ||| |||| |||||||||||||||||||||||| | |||||||  | ||||||||||||||
Db  8977 CATTAGTAtGtTTATGCTTCACTCCTAGGAGAAGTACGACGAGCGTCGTCGTCAAGAGAACAAGC  8918

Qy   742 attaaaattgtgactcctgactggttctggatgtgcgtatcagagaaaaaccaaaaaggac  801
         |||| | || |||||| ||||||| | |||| | || | | | ||||||| |||||||||
Db  8917 ATTAAGATTGTTACCCCTGACTGGGTCCTCGACTGTGTGTCTGAGAGGAGAGGAGAAGGAT  8858

Qy   802 gaagcgatttatcatcccgtctgattatttattatgaagaggaagaggaaggagaggagaggagg  861
         |||| ||| ||| | || ||||| ||| || |||||||||||||| |||| ||||| |||||
Db  8857 GAAGCTCTTACCATCCACGTCTGATCCATCTATGAGGAGGAA------GAGGAGGAGGAG  8804

Qy   862 gagaggaaggtagaaaatgaggaacaagaatctcaqaatgaggggtagtacagatgaggag  921
         ||| |||| || |||| ||||| |||| ||| | || || ||| |||| || |||||||||
Db  8803 GAGGAAGAGGCAGACAACAGACAGCAGAGAGGCTTCAGAACGACGAGGGTAGCAC---TGACAAG  8747

Qy   922 tcaagccctgccagctctcaaggaagggtcccttcaaggtgaccaacagtttcacctaaa  981
         ||||| || ||||||||| ||||| |||| || || |||| |||| | ||||||||
Db  8746 TCCAGTGCGAGCCAGCTCTGCGGAGCCCCTCCCCT---GCTGACACAGCGGTGCTCTCCCAAG  8690

Qy   982 tccaacactgaaaaatctaaagggggaattaatgttgatgattcttcagattccatcaccg  1041
         ||| ||||| |||| |||||| ||||||| ||| |||| || |||||| ||||| |||
Db  8689 GCCAGGGCTGAGGTAtCCAAAGGGAGtGTTGATGTTTGAtGAGCTCTTCTGACTCATCCCT  8630

Qy  1042 gaaaacaacagggacgagaaatttaactgaccccgggcggaagtcccacagttagctgcagca  1101
         ||||||| ||| || ||| |||||||| ||| |||| | |||||||||| ||| ||||||
Db  8629 GAAAACAGGAGACGGCAGCTTGAACTCGCCCCCGTCGAAGCCCCCACTCAGCACAGCC  8570

Qy  1102 aaacgcagcqtgcctcaggaaagggagcctgggttgttgattaacttgtgtgccaatgtcccca  1161
         ||| |||| ||||||||||| |||| ||||| ||| |||||||||| |||||||||| |||||
Db  8569 CAGCACAGGCGCTGCTCAGGAAAAGGACCGGGACTCATTAACCTGTGTGCCAAATGTCCCA  8510

Qy  1162 cccgtcccagtaacaatttgcccccctgaggtccgggggtaattteaatggctgctggacaa  1221
         || |||||||| || | ||||| || ||| || ||||||| || || |||||| |||||| |
Db  8509 CCTGTCCCAGGGGACNCATTCTGCCTCTTGACNATCGGGGGTNAACTTGANTGGCCGCTGGACNA  8450

Qy  1222 aacctccaaagttctgaaagatcaqaaatgatagactacctgagtccagctgtacggaca  1281
         ||||||| |||| ||| | || | ||||| | || ||||||||| ||| |||| |||||
Db  8449 AACCTCCAGAGTGTCNGACGGCGCCGAAATCGTTACGGACACTGAGCCACTCTCCGGACG  8390

Qy  1282 claqggaattattaalaactgacattcaqcagcatgaaccggcccatcaaaatgtagca  1341
         | ||||| ||| || | |||| |||| ||||||||| |||| |||||| |||| |||
Db  8389 TTCGGAAACATTACNAACAACGCTGATATNCCCAGGCAAATCAGCCGGCCANTGTTGCA  8330

Qy  1342 catat 1346
         ||| |
Db  8329 CATGT 8325
```

LOCUS   AC074040   178419 bp   DNA   linear   HTG 13-JUL-2000
DEFINITION Mus musculus chromosome 5 clone RP23-130M12 strain C57BL6/J,
           WORKING DRAFT SEQUENCE, 24 unordered pieces.
ACCESSION   AC074040
VERSION   AC074040.1 GI:9087212
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   house mouse.
  ORGANISM Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 178419)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
           Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
           Huang,M.C., Idol,J., Lee-Lin,S.-Q., Maduro,Q.L., Maduro,V.B.,
           Mastrian,S.D., McCloskey,J.C., Ojodu,M.A., Pearson,R.,
           Stanlripop,S., Summers,T.J., Thomas,J., Thomas,P.J.,
           Tsongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
           Wetherby,K.D. and Green,E.D.
  TITLE   NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 178419)
  AUTHORS   Green,E.D.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-JUL-2000) NIH Intramural Sequencing Center,
           NISC Mouse Sequencing Initiative,
           Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT

```
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information ----------------
            Center project name: tk
            Center clone name: 130M12
            ------------------- Summary Statistics ----------------
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 164732 bases at least Q40
            Consensus quality: 168102 bases at least Q30
            Consensus quality: 170232 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 177000; pulse-field-gel
            Insert size: 176119; sum-of-contigs
            Quality coverage: 4.22x in Q20 bases; agarose-fp
            Quality coverage: 4.37x in Q20 bases; pulse-field-gel
            Quality coverage: 4.39x in Q20 bases; sum-of-contigs
            --------------------------------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 24 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          *        1  2335: contig of 2335 bp in length
          *     2336  2435: gap of unknown length
          *     2436  4708: contig of 2273 bp in length
          *     4709  4808: gap of unknown length
          *     4809  7741: contig of 2933 bp in length
          *     7742  7841: gap of unknown length
          *     7842 11301: contig of 3460 bp in length
          *    11302 11401: gap of unknown length
          *    11402 16198: contig of 4797 bp in length
          *    16199 16298: gap of unknown length
          *    16299 21111: contig of 4813 bp in length
          *    21112 21211: gap of unknown length
          *    21212 26078: contig of 4867 bp in length
          *    26079 26178: gap of unknown length
          *    26179 31586: contig of 5408 bp in length
          *    31587 31686: gap of unknown length
          *    31687 38653: contig of 6967 bp in length
          *    38654 38753: gap of unknown length
          *    38754 44405: contig of 5652 bp in length
          *    44406 44505: gap of unknown length
          *    44506 50909: contig of 6404 bp in length
          *    50910 51009: gap of unknown length
          *    51010 56618: contig of 5609 bp in length
          *    56619 56718: gap of unknown length
          *    56719 65690: contig of 8972 bp in length
          *    65691 65790: gap of unknown length
          *    65791 75015: contig of 9225 bp in length
          *    75016 75115: gap of unknown length
          *    75116 82226: contig of 7011 bp in length
          *    82227 82326: gap of unknown length
          *    82327 89936: contig of 7710 bp in length
          *    89937 90036: gap of unknown length
          *    90037 99253: contig of 9217 bp in length
```

```
*    99254    99353: gap of unknown length
*    99354   110325: contig of 10972 bp in length
*   110325   110425: gap of unknown length
*   110426   119465: contig of 9040 bp in length
*   119466   119565: gap of unknown length
*   119566   128843: contig of 9278 bp in length
*   128844   128943: gap of unknown length
*   128944   138762: contig of 9819 bp in length
*   138763   138862: gap of unknown length
*   138863   152091: contig of 13229 bp in length
*   152092   152191: gap of unknown length
*   152192   164766: contig of 12575 bp in length
*   164767   164866: gap of unknown length
*   164867   178419: contig of 13553 bp in length.

FEATURES             Location/Qualifiers
     source          1..178419
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-130M12"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1..2335
                     /note="assembly_fragment"
     misc_feature    2436..4708
                     /note="assembly_fragment"
     misc_feature    4809..7741
                     /note="assembly_fragment"
     misc_feature    7842..11301
                     /note="assembly_fragment"
     misc_feature    11402..16698
                     /note="assembly_fragment"
     misc_feature    16299..21111
                     /note="assembly_fragment"
     misc_feature    21212..26078
                     /note="assembly_fragment"
     misc_feature    26179..31586
                     /note="assembly_fragment"
     misc_feature    31687..38653
                     /note="assembly_fragment"
     misc_feature    38754..44405
                     /note="assembly_fragment"
     misc_feature    44506..50909
                     /note="assembly_fragment"
     misc_feature    51010..56618
                     /note="assembly_fragment"
     misc_feature    56719..65690
                     /note="assembly_fragment"
     misc_feature    65791..75015
                     /note="assembly_fragment"
     misc_feature    75116..82126
                     /note="assembly_fragment"
     misc_feature    82227..89936
                     /note="assembly_fragment
                     vector_side:left"
                     clone_end:SP6
     misc_feature    9037..99253
                     /note="assembly_fragment"
     misc_feature    99354..110325
                     /note="assembly_fragment"
     misc_feature    110426..119465
                     /note="assembly_fragment"
     misc_feature    119566..128843
                     /note="assembly_fragment"
     misc_feature    128944..138762
                     /note="assembly_fragment"
     misc_feature    138863..152091
                     /note="assembly_fragment"
     misc_feature    152192..164766
                     /note="assembly_fragment"
     misc_feature    164867..178419
                     /note="assembly_fragment"
BASE COUNT    48439 a   40696 c   40698 g   46277 t   2309 others
```

```
RESULT 11
LOCUS       AC079217              177205 bp    DNA     linear   HTG 24-AUG-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-381D23 strain C57BL6/J,
            WORKING DRAFT SEQUENCE, 20 unordered pieces.
ACCESSION   AC079217
VERSION     AC079217.1  GI:9887697
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 177205)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Idol,J., Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L.,
            Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Stantripop,S., Summers,T.J., Thomas,J.W.,
            Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
            Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
```

```
ORIGIN
Query Match   9.7%;  Score 348.6;  DB 2;  Length 178419;
Best Local Similarity  75.1%;  Pred. No. 2.7e-65;
Matches  480;  Conservative  0;  Mismatches 144;  Indels 15;  Gaps 3;

Qy    708 ggggaaatacgaatgtgcttaaagtcggagcaagtattaaaatttgactcctgactggt 767
          |||||||| ||||||| |||||||||||||||||||||||||||||||| |||||||||
Db  97839 GGAGAAATACGAGCGTGCGTGTCAAAAGAACAAGTATTAAGATTGTCACCCGGACTGGGT 97898

Qy    768 tctggattcgtatcacagagaaaaccaaaaagaccgaagcaattttatcctcgtctgat 827
          |||| ||| |||||||||||| |||||||||||||||||||||||| ||| || |||||
Db  97899 CCTGGACTGTGTTGCGTCTGAGGAGGAGGAGGAGGATGAAACCGTTTTACCATCCACCTGAT 97958

Qy    828 tattatgagaggaagaggaggaagaggaggaagaggaggaagaggaggaaca 887
          ||||  |||||  ||  |||||  |||||||  ||  |||  |||||  ||
Db  97959 CATCTATGAGGAGGAGGAGGAAGAGGAGGAGGAAGAG------GGAGCAACGAGGAGCA 98009

Qy    888 agattctcaagaatgagggttagtacagatgaggaagtcaagaaccctgccagctctcaagagg 947
          |||||| ||||  ||||||  |  ||||  |||||| ||  ||    |||||||| ||||| |||
Db  98010 GGACTCCCAGAACGAGGGGCAGCAC---TGAGAAGTCCAGTGTCGCCAGCTTCTGCAGTGGC 98066

Qy    948 gtctcctcaagtgaccagcagtttcacctaaatccaacactgaaaaatctaaagggga 1007
          |  |   ||  |||||||| ||  |  |      ||    |||||||||||||| ||
Db  98067 CTCCCCT---CCTGACCAGCCGGTGCCTCTCCCCAAGCCCACGGGGTGAG../UTCCAAAGGAGA 98123

Qy   1008 attaaatgtttgatgattcttcagaatcatcatcaccggaaaaaacaggagagaaatttaaactg 1067
          ||| ||||| ||||| ||  | ||  ||||| ||      ||||  ||||| |  |    |||||||
Db  98124 GTTGCATGTTGTAGTGACTCNTTCTGACTCATCCCGTCNAAAAGCAGGACGCGACGGCTTGAACTG 98183

Qy   1068 gacccgccgccgaagtcccacagttagctgagcaaaacgcaggctgcctcaggagaaga 1127
          | |||| |||| |||  ||||   |  |  ||||  |||  ||||| |||| |||| || 
Db  98184 GGCCCGCTGCTGAAGCCCCACCACTCAACACAGCCCGCCGAGCTGCGCCTCCTCAGGGGAAAGG 98243

Qy   1128 gcctgggttgattaacttgtgtgccaaatgtcccacccgtcccaggtaacaatttgccccc 1187
          ||| ||  | |||||| |||||||||||||||||||| ||  ||| ||  ||||  ||| ||
Db  98244 GCCAGG../TCATTAACCTGTGTGCCAAATGTCCCACCTGTTCCAGCGGGCACATTYCTGCCGTCC 98303

Qy   1188 tgaggtccgggggtaatttaatgctgctggacadaaacctccaaagttctgaaagatcaga 1247
          ||| | ||| |||| || ||  || | |||||| ||||||||| ||| | ||| ||||||||
Db  98304 TGACATGCGGCGGTAACTTGATGGCCCCCTGGACAAAACCTCCAGAATTYCGAGCGGTCAGA 98363

Qy   1248 aatgatagctacctgggagtccagctgtagggaacctgaggaatattactaatctgctga 1307
          ||| | |||  |||||| |   ||||||||  ||  | | |||| |||||| | |||||||
Db  98364 AATACTAGGGACCTGGAGCCCAGCGTGCCAGCTGTTGCGAAAACATTAACAAAACAAATGCTGA 98423

Qy   1308 cattcagacagatgaaccggccatcaaaatgtagcacatat 1346
          |||| |||||||| || ||||||||| |||||| ||||  |
Db  98424 TATCCACCACAGATCAACCGGCCATCCAAATGTTGCACAVGT 98462
```

JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 172205)
AUTHORS   Green,E.D.
TITLE   Direct Submission
JOURNAL   Submitted (24-AUG-2000) NIH Intramural Sequencing Center, 8717
           Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT   Center: NIH Intramural Sequencing Center
           Center code: NISC
           Web site: http://www.nisc.nih.gov
           Contact: nisc_mouse@nhgri.nih.gov
           -------------- Project Information
           Center project name: t1
           Center clone name: 381D23
           -------------- Summary Statistics
           Sequencing vector: plasmid; n/a; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.990319
           Consensus quality: 164801 bases at least Q40
           Consensus quality: 167308 bases at least Q30
           Consensus quality: 168762 bases at least Q20
           Insert size: 209000; agarose-fp
           Insert size: 188000; pulse-field-gel
           Insert size: 175305; sum-of-contigs
           Quality coverage: 4.10x in Q20 bases; agarose-fp
           Quality coverage: 4.56x in Q20 bases; pulse-field-gel
           Quality coverage: 4.89x in Q20 bases; sum-of-contigs
           ----

   * NOTE: This is a 'working draft' sequence. It currently
   * consists of 20 contigs. The true order of the pieces
   * is not known and their order in this sequence record is
   * arbitrary. Gaps between the contigs are represented as
   * runs of N, but the exact sizes of the gaps are unknown.
   * This record will be updated with the finished sequence
   * as soon as it is available and the accession number will
   * be preserved.
   *
   *   1   3158: contig of 3158 bp in length
   *   3159   3258: gap of unknown length
   *   3259   6522: contig of 3264 bp in length
   *   6523   6622: gap of unknown length
   *   6623   10277: contig of 3655 bp in length
   *   10278   10377: gap of unknown length
   *   10378   14299: contig of 3922 bp in length
   *   14300   14399: gap of unknown length
   *   14400   18285: contig of 3886 bp in length
   *   18286   18385: gap of unknown length
   *   18386   24955: contig of 6570 bp in length
   *   24956   25055: gap of unknown length
   *   25056   30203: contig of 5148 bp in length
   *   30204   30303: gap of unknown length
   *   30304   36272: contig of 5969 bp in length
   *   36273   36372: gap of unknown length
   *   36373   40990: contig of 4618 bp in length
   *   40991   41090: gap of unknown length
   *   41091   49862: contig of 8772 bp in length
   *   49863   49962: gap of unknown length
   *   49963   55586: contig of 5624 bp in length
   *   55587   55686: gap of unknown length
   *   55687   65769: contig of 10083 bp in length
   *   65770   65869: gap of unknown length
   *   65870   76152: contig of 10283 bp in length
   *   76153   76252: gap of unknown length
   *   76253   87685: contig of 11433 bp in length
   *   87686   87785: gap of unknown length
   *   87786   97230: contig of 9445 bp in length
   *   97231   97330: gap of unknown length
   *   97331 112484: contig of 15154 bp in length
   * 112485 112584: gap of unknown length
   * 112585 127931: contig of 15347 bp in length
   * 127932 128031: gap of unknown length
   * 128032 143545: contig of 15514 bp in length
   * 143546 143645: gap of unknown length
   * 143646 159112: contig of 15467 bp in length

   *   1,9913   159212: gap of unknown length
   *   159213   177205: 177205: contig of 17993 bp in length.
FEATURES   Location/Qualifiers
   source   1..172205
                   /organism="Mus musculus"
                   /strain="C57BL6/J"
                   /db_xref="taxon:10090"
                   /chromosome="5"
                   /clone="RP23-381D23"
                   /clone_lib="RPCI mouse BAC library 23"
   misc_feature   1..3158
                   /note="assembly_fragment"
   misc_feature   3259..6522
                   /note="assembly_fragment"
   misc_feature   6623..10277
                   /note="assembly_fragment"
   misc_feature   10378..14299
                   /note="assembly_fragment"
   misc_feature   14400..18285
                   /note="assembly_fragment"
   misc_feature   18386..24955
                   /note="assembly_fragment"
   misc_feature   25056..30203
                   /note="assembly_fragment"
   misc_feature   30304..36272
                   /note="assembly_fragment"
   misc_feature   36373..40990
                   /note="assembly_fragment"
   misc_feature   41091..49862
                   /note="assembly_fragment"
   misc_feature   49963..55586
                   /note="assembly_fragment"
   misc_feature   55687..65769
                   clone_end::SP6
                   vector_side::right"
   misc_feature   65870..76152
                   /note="assembly_fragment"
   misc_feature   76253..87685
                   /note="assembly_fragment"
   misc_feature   87786..97230
                   /note="assembly_fragment"
   misc_feature   97331..112484
                   /note="assembly_fragment"
   misc_feature   112585..127931
                   /note="assembly_fragment"
   misc_feature   128032..143545
                   /note="assembly_fragment"
   misc_feature   143646..159112
                   /note="assembly_fragment"
   misc_feature   159213..177205
                   /note="assembly_fragment"
BASE COUNT   49012 a 39764 c 38186 g 48267 t   1976 others
ORIGIN

Query Match   9.3%; Score 334.4; DB 2; Length 177205;
Best Local Similarity   74.7%; Pred. No. 3.4e-62;
Matches 478; Conservative   0; Mismatches 146; Indels 16; Gaps   4;

Qy   708 gggaaatacgaatgtgcttaagcgagcaagtattaaattgtgactcctgactgggt 767
          ||| || ||||||||| |||||| | |||||| | |||||| || ||| |||||||
Db 73083 GGAGAAAATACGAGCGTGCTGTCAAAACAAGAACAAGTATTAAGATTGTCACCCCGGACTGGGG 73142

Qy   768 t-ctggattggtgtatcagagaaaacaaaaaggcgaagcattttatcatcctcgtctga 826
          ||||| || || || |||| ||||| || |||| ||||| ||| | | |||| | |||| |
Db 73143 CCCTGAACTGTGTGTCTGCGAAGAGGAGGATGAAGCGTTTTACCATCCACGCCTGA 73202

Qy   827 ttattatgaagaggtagtacagatagagaagtcaagccctgccagctctcaagaag 886
          | | | | |||| |||| ||| ||| |||| ||| |||| | |||| | |||| |
Db 73203 TCATCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG------GGAGGACAACGAGGAGC 73253

Qy   887 aagaatctcagaatgagggtagtacagatagagaagtcaagccctgccagctctcaagaag 946

RESULT 12

LOCUS   HSU23862   633 bp   mRNA   linear   PRI 24-JUL-2001
DEFINITION Human clone mcag32 chromosome 7 CTG repeat region.
ACCESSION   U23862
VERSION   U23862.1 GI:777403
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 633)
  AUTHORS   Yamagata,K., Takeda,J., Menzel,S., Chen,X., Eng,S., Lim,L.R.,
            Concannon,P., Hanis,C.L., Spielman,R.S., Cox,N.J. and Bell,G.I.
  TITLE   Searching for triplet repeats expressed in skeletal muscle
  JOURNAL   Diabetologia 39 (6), 725-730 (1996)
  MEDLINE   96375458
  PUBMED   8781769
REFERENCE   2 (bases 1 to 633)
  AUTHORS   Bell,G.I.
  TITLE   Direct Submission
  JOURNAL   Submitted (31-MAR-1995) Graeme I. Bell, Howard Hughes Medical
            Institute, University of Chicago, 5841 South Maryland Avenue,
            Chicago, IL 60637, USA
FEATURES   Location/Qualifiers
     source   1..633
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /clone="mcag32"
                     /sex="female"
                     /tissue_type="skeletal muscle"
                     /clone_lib="Stratagene catalogue no. 936215"
                     /dev_stage="adult"
     repeat_region   356..433
                     /rpt_family="CTG"
BASE COUNT   107 a   152 c   207 g   167 t
ORIGIN

Query Match   8.6%; Score 308.2; DB 9; Length 633;
Best Local Similarity 82.0%; Pred. No. 7.9e-57;
Matches 355; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy   1465 cacatgctacagcagcagcaggccagacggcagcagcagcagcagcaccggtttacac 1524
          || ||||||||||||||| |||||||||||| ||||||||||||||||||||||||| |
Db   433 CACCTCCAGCAGCAGCACCAGCAGCAGCAGCAGTTCAGCAGCAGCAGCTCCAGCAGCACCAGGCCCAGGCCCCAGGGCCCCAGGCCCCAGGCCCCAGGCCCCAGTGCACCAG 374

Qy   1525 cttcagcccccagcagcagataatgcagctccagcagcagcagcagcagcatctcagcaa 1584
          |||||||| ||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   373 CAGCAGCAGCAGCAGCAGCAGGATGCAAAGTCAGACAGCGCCACACTTGAGTCAGACGTCACAG 314

Qy   1585 ccttaccccagcagcgcccgccgaatccatttcagcagcagcagcagcagcaacagcagcca 1644
          ||||| ||||||| ||||| || || || |||||||||||||||||||||| ||| || ||
Db   313 GCGCTGCAGCATCAGCAGTTCCACCTCAGCAGCACCAGCAGCAGCAACAGCAGCAGCCA 254

Qy   1645 ccaccatcgcctccagcacatcagcttttggacatgatccagcagtggagattccagaa 1704
          |||||||| ||| ||||||||| || |||||||||| || || || ||||||||||||
Db   253 CCACCATCGCCTCCAGCAGCACATCAGCTTTTGGACATGATCCAGCAGCACTGGAGATTCCAGAA 194

Qy   1705 gaaggcttcttattggatgtgtcgtttgcaaatgcggagttatccagcagcagatgtctgat 1764
          |||||||||||| ||||||| |||||||||| ||||||||||||||||||||||||||||
Db   193 GAAGGCTTCTTATTGGATGTGTCGTTTGCAAATGCGGAGTTATCCAGCAGCAGATGTCTGAT 134

Qy   1765 aagcaactgtggtcacctgagaaaggatataaatccaggccacatggcggcaactgttgaccc 1824
          ||||||||||||||| ||| |||||| |||||||||||||||||||||||||||||||||||
Db   133 AAGCAACTGTGGTCGGCCACCTGGGAAAAGGATAATAATCCAGGGCACTGTTGACCCC 74

Qy   1825 accttcacgagtcagatcgacgcaacctctctgtgagagtcaaagtcaagcagcgcgtatgca 1884
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 ACCTTCACGAGTCGGATCGACGCACCGCACCTTCTCTCTGCAGAGTCAAGTCAGCAGCAGCGCGTATGCA 14

Qy   1885 caggcaataagag 1897
          |||||||||||||
Db   13 CAGGCAATAAGAG 1

RESULT 13

LOCUS   G15998   687 bp   DNA   linear   STS 19-JAN-1996
DEFINITION human STS CHLC.GCT8H06.P11207 clone GCT8H06, sequence tagged site.
ACCESSION   G15998
VERSION   G15998.1 GI:1161887
KEYWORDS   STS; STS sequence; primer; sequence tagged site.
SOURCE   human vector-pJCP1 host-E.coli dut-ung+ (DH10B) Marker Selected
            genomic DNA prepared from XY individual of French nationality.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 687)
  AUTHORS   Murray,J., Sheffield,V., Weber,J.L., Duyk,G. and Buetow,K.H.
  TITLE   Cooperative Human Linkage Center
  JOURNAL   Unpublished
COMMENT   Synonyms: GCT8H06, CHLC.GCT8H06.T11206
            Uofl
            The University of Iowa
            Department of Pediatrics, Iowa City, IA 52242, USA
            Tel: (319) 356-3508
            Fax: (319) 356-3347
            Email: jeff-murray@uiowa.edu
            Contact: Dr. Jeffrey C. Murray
            Primer A: CCCCACACTCAAACACTCTC
            Primer B: GACACACATGCTACAGCAGC
            STS size: 347
            PCR Profile:

Protocol:

Template: 30ng genomic DNA
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C

Db   73254 AGGACTCCCAGCAGAACCAGCAGCCAGCAC---TGAGAAGTCCAGTCTGTCCCCAGCTCTGCCAGTCG 73310
Qy   947 ggtctccttcaggtgaccagcagctttcacctaatccaacactgaaaaatctaaagggg 1006
Db   73311 CCTCCCCTT---GCTCAGCAGCCGTTGCTCTCCCTTCCAAGCCCTTGCAGGTGTTCCAAAGGAG 73367
Qy   1007 aattaatgtttgatgtattcttcagatttcatcaccgagacaacagagagaaatttaaact 1066
Db   73368 AGTTGAGTTGATGACTCTTCTGACTCATCCCCTGAAAAAGCAGAGCAGCCAGCTTGAACT 73427
Qy   1067 ggacccggcccggaagtcccagttagctgcgcaaaagcaggctgcctcaggggaagg 1126
Db   73428 GGGCCCCCTGCTGAAGCCCCACCACTCAACAAGCACACAGCCCACCGCAGCTGCCTCAGGGAAAAAG 73487
Qy   1127 agcctgggttgattaacttgtgtgccaatgtcccaaccgtcccaggtaacatttgccc 1186
Db   73488 GCCCAGGACTCATTAACCTGTGTCCCAAGTCCCCGTTCCAGGGGACATTCTCCTC 73547
Qy   1187 ctgaggtccgggtgaattaatggctgctgctgggacaaaacctccaaagttctgaagataca 1246
Db   73548 CTGACATGCGGGGTAACTTGATGGCCCCTGGACAAAAACCTCCAGAATTCTGAGCGGTCAG 73607
Qy   1247 aaatagataactacctggtatccagcctgtacggacactgaggaatattactaataataatgctg 1306
Db   73608 AAATACTAGGGACCTGGAGCCCAGCTGTCCGGACGTTGCCAAACAATTACAAACAATGCTG 73667
Qy   1307 acattcagcagatgaaccgccatcaaatgtagcacatat 1346
Db   73668 ATATCCAGCAGAATCAACCGGCCATCCATGTTGCACATGT 73707

```
RESULT 14
LOCUS       AX186240
AX186240
LOCUS       AX186240            690 bp    DNA    linear   PAT 06-AUG-2001
DEFINITION  Sequence 1935 from Patent WO0142467.
ACCESSION   AX186240
VERSION     AX186240.1  GI:15137668
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 690)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
```

FEATURES   Location/Qualifiers
     source   1..690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT   167 a   148 c   164 g   203 t   5 others
ORIGIN

Query Match   79.9%; Score 263.4; DB 11; Length 687;
Best Local Similarity   7.4%; Pred. No. 5.1e-47;
Matches 345; Conservative   0; Mismatches   56; Indels   31; Gaps   2;

Qy   1318 atgaaccggccatccaaatgtagcacatatcttcagacactacgaaaaat 1377
                |||| |||||||||||||||||||| |||||||||||||||||||||||
Db   616 ATGTTTCACCTAACTTTGTTTTGTAGATCTTACAGACTCTTTCAGNACCTACGAAAAT 557

Qy   1378 ttagaaacagcaggtgaatcaacagccagcaggaggacaacaaaatgccaatgctgtgtt 1437
                ||| |||||||||||||| || ||||||||||||||||||||||||| | ||||||
Db   556 TTAGNCAGNAGCNAGGTCNAATCANCAGCCAGCAGGACAATCCAATCCAGTGCTGTTT 497

Qy   1438 agccaagtgaaagtgactccagagacacanatgctacagcagcagcagggcccaagcag 1497
                ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   496 AGCCAAGTGAAAGTGACTCCAGAGACACAnTCTCAGAGACACAACCACCAGCCCCAGCAC 437

Qy   1498 cagcagcagcagcagcacccggtttacacctccagccccagcagagatgagctccagcag 1557
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   436 CAGCAGCAGCAGCACCCGGTTTACACCTTCAGCCCCAGCAGNTANTGCGAGCTCCAGCAG 377

Qy   1558 cagcagcagcagcaga----------------------tctctcagcaaacctaccc 1593
                ||||||||||||||| ||||| ||||||||||||
Db   376 CAGCAGCAGCAGCAGACGGCAAACTCAGACAGCCCACACTTGAGTCAGACGTCACAGCCG 317

Qy   1594 cagcagcagcgccgcatccatttca----------cagcagcagcagcaacagcagccacc 1646
                |||||| ||| | ||||| ||| |||||||||||||||||||||||||||
Db   316 CTGCAGCCATCAGGTTCCACCCTCAGCCAGCCCCGCAGCAGCAGCAGCAACACCAGCCACC 257

Qy   1647 accatgcgcctccagcaatcagcattcttggacatgatccagagtggagattccaagaga 1706
                || |||||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db   256 ACCATCGCCTCAGCAGCATCAGCTTTTTGGACANTGATCCAGCAGTNGGAGAGTTAGTAATG 197

Qy   1707 aggcttcttatt 1718
                ||| || || ||
Db   196 GAGATTATTTTT 185
```

```
BASE COUNT      218 a    172 g    124 t        2 others
ORIGIN

Query Match     95.9%;  Score 204.6;  DB 6;  Length 690;
Best Local Similarity    5.7%;  Pred. No. 3.7e-34;
Matches  210;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     1073 cggccgaagtcccacagttagctgcagcaaaacgcaggctgcctccagggaaaggagcctg 1132
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       23 CGGCCGGAAGTCCCACAGTTAGCTCGACAGAAAAACGCAGGGTGCCTCCAGGGAAAGGAGCCTG 82

Qy     1133 ggttgattaacttgtcttgtccaagtgtcccacaccgtcccaggtaacactttgccccctgagg 1192
                ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       83 GGTTGATTAACTTGTCTTGTCCAATGTCCCACACCGTCCCAGGTAACATTTTGCCCCCTGAGG 142

Qy     1193 tccggggtaatttaatggctgtggtggacaaaaacctccaaagtttctgaaaagatcagaaatga 1252
                |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db      143 TCCGGGGTAATTTAATGGCTGTGGTGGACAAAAACCTCCAAAGTTCTGAAAGATTCAGAAATGA 202

Qy     1253 tagctacctggagtccagctgtacggacactgaggaata 1291
                |||||||||||||||||||||||||||||||||| | |
Db      203 TAGCTACCTGGAGTCCAGCTGTACGGACACTGGCCGTTAA 241
```

```
JOURNAL     Patent: WO 0142467-A 1935 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..690
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      218 a    172 g    124 t        2 others
```

```
RESULT 15
LOCUS       AX187445
AX187445
LOCUS       AX187445           504 bp    DNA    linear   PAT 06-AUG-2001
DEFINITION  Sequence 3140 from Patent WO0142467.
ACCESSION   AX187445
VERSION     AX187445.1  GI:15138906
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   Schlegel,R., Deeds,J., Berger,A. and Zhao,X.
  TITLE     Genes, compositions, kits, and methods for identification,
            assessment, prevention, and therapy of cervical cancer
  JOURNAL   Patent: WO 0142467-A 3140 14-JUN-2001;
            Millennium Predictive Medicine, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..504
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      147 a    135 c    121 g    101 t
ORIGIN

Query Match     94.0%;  Score 186.8;  DB 6;  Length 504;
Best Local Similarity    5.2%;  Pred. No. 2.8e-30;
Matches  205;  Conservative    0;  Mismatches   12;  Indels    1;  Gaps    1;

Qy     1075 gccgaagtcccacagttagctgcagcaaaacgcaggctgcctccagggaaggagcctggg 1134
                |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       21 GGCCGAGTCCCACAGTTAGCTGCAGCAAAAACGCAGGGTGCCTCACGGGAAAGGAGCCTGGG 80

Qy     1135 ttgattaacttgtgtgccaagtccaccccgtccacccagaggtaacattttgcccctgaggtc 1194
                |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db       81 TTGATTAACTTGTGTGTCAATGTCCACCCGTCCCACCCAGGTAACATTTTGCCCCTGAGGTC 140

Qy     1195 cggggtaatttaatggctgtggacaaaaacctccaaagttc-tgaaagatcagaaatgat 1253
                ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db      141 CGGGGTAATTTAATGGCTGTGGACAAAAACCTCCAAAGTTCTTGAAAGATCAGAAATGAT 200

Qy     1254 agctacctggagtccagctgtacgacactgaggaata 1291
```

Db   201 AGCTACCTGGAGTCCAGCTGTACGGCACTTGGCGTAAA 238
            ||||||||||||||||||||||||||| || | | |

Search completed: June 4, 2002, 15:53:06
Job time: 17024 sec

OM nucleic - nucleic search, using sw model

```
                    GenCore version 4.5
           Copyright (c) 1993 - 2000 Compugen Ltd.
```

Run on: June 4, 2002, 11:09:22 ; Search time 3648.51 Seconds
                                         (without alignments)
                                         13243.516 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacggct..........gaacttagaaaaaaaaaa 3580

Scoring table: IDENTIFY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Minimum DH seq length: 0
Maximum DB seq length: 2000000000

Database :   EST:*

```
  1: em_estba:*
  2: em_esthum:*
  3: em_estin:*
  4: em_estmu:*
  5: em_estov:*
  6: em_estpl:*
  7: em_estro:*
  8: em_htc:*
  9: gb_est1:*
 10: gb_est2:*
 11: gb_gss:*
 12: gb_htc:*
 13: em_gss_hum:*
 14: em_gss_inv:*
 15: em_gss_pln:*
 16: em_gss_vrt:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	Query Match	Length	DB	ID	Description
1	c	809.4	22.6	1059	10	BM470958	BM470958 AGENCOURT
2	c	707.2	19.8	760	9	AI671887	AI671887 wb41c03.x
3	c	700.6	19.6	780	9	AW073165	AW073165 wy94g07.x
4	c	698.2	19.5	811	9	AI357401	AI357401 qy13d06.x
5	c	653.2	18.2	759	10	BE260052	BE260052 601150905
6	c	650.4	18.2	652	9	AII33662	AII33662 HA2284 Hu
7	c	639.8	17.9	680	9	AI741935	AI741935 wg29g05.x
8	c	583	16.3	669	10	BG900296	BG900296 HOA37-1-B
9	c	581.2	16.2	625	9	AU123803	AU123803 AU123803
10	c	573.2	16.0	803	10	BE541477	BE541477 601067929
11	c	558.4	15.6	597	9	AW009159	AW009159 ws78b09.x
12	c	547.2	15.3	552	9	AW004836	AW004836 ws93c05.x
13	c	542	15.1	598	9	BF939756	BF939756 nac83a11.
14	c	540.4	15.1	598	9	AI807116	AI807116 we85g11.x
15	c	539.6	15.1	551	9	AI393049	AI393049 tg25c09.x
16	c	528.6	14.8	541	9	AA846820	AA846820 aj99g08.s
17	c	518.8	14.5	535	9	AW937777	AW937777 QV3-DT004

ALIGNMENTS

18		507.2	14.2	1154	10	BM461168 BM461168 AGENCOURT
19		500.8	14.0	504	10	BF876036 BF876036 IL3-ET011
20		499	13.9	516	9	AW390801 AW390801 RC6-ST019
21		499	13.9	531	9	AL135690 AL135690 DKFZp762p
22		496	13.9	730	10	BI248290 BI248290 60296688.3
23	c	485.8	13.6	677	10	BG259569 BG259569 60237870.2
24	c	473	13.2	514	9	AI625826 AI625826 ty55b06.x
25	c	472	13.2	472	9	AI362238 AI362238 qy45b09.x
26	c	471.8	13.2	520	9	AW341490 AW341490 hd10d04.x
27	c	466.8	13.1	504	9	AA587951 AA587951 nm88g02.s
28	c	466.8	13.0	480	9	AW303585 AW303585 xv20h08.x
29	c	463.8	13.0	540	10	BF916607 BF916607 IL3-UT011
30	c	454.4	12.7	480	10	N51347 N51347 yz15g08.s1
31	c	445	12.4	445	10	BE67288 BE677288 7d83ac.x
32	c	441.2	12.3	457	9	AL042281 AL042281 DKFZp434.
33	c	439.4	12.3	888	10	BE573124 BE573124 601331567
34		438	12.2	872	10	BI735116 BI735116 603356256
35	c	434.6	12.1	640	10	BI401433 BI401433 MI-P-CP0-
36	c	433.4	12.1	435	10	AI080755 AI080755 ox54h03.x
37	c	432.4	12.1	547	9	AW867972 AW867972 RC0-SN004
38	c	425.2	11.9	430	9	AW117548 AW117548 xd93b10.x
39	c	424.4	11.9	426	9	AW663176 AW663176 hh75f01.y
40	c	416.4	11.9	419	9	AW183769 AW183769 xj87f05.x
41	c	416.4	11.6	583	10	BF320169 BF320169 u251c09.y
42	c	412.4	11.5	738	10	BI684354 BI684354 603308191
43	c	412.2	11.5	425	9	AA917929 AA917929 ol67q01.s
44	c	411.2	11.5	506	10	N51997 N51997 yz07h04.s1
45	c	411.2	11.5	555	10	BF012553 BF012553 ux57g08.y
```

RESULT 1
BM470958

LOCUS   BM470958   1059 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION AGENCOURT_6462994 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5533689
            5', mRNA sequence.
ACCESSION   BM470958
VERSION   BM470958.1 GI:18520000
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1059)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMI2218 row: p column: 10
            High quality sequence stop: 751.
FEATURES   Location/Qualifiers
     source   1..1059
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5533689"
                     /clone_lib="NIH_MGC_71"
                     /tissue_type="leiomyosarcoma"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     /lab_host="DH10B (phage-resistant)"
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT."
BASE COUNT   307 a   245 c   242 g   260 t   5 others
ORIGIN
```

RESULT 2
AI671887/c

```
Query Match    22.6%;  Score 809.4;  DB 10;  Length 1059;
Best Local Similarity  97.3%;  Pred. No. 4.1e-124;
Matches 896;  Conservative  0;  Mismatches  18;  Indels  7;  Gaps  7;
```

```
Qy   1747 ccagaacacagatgtctgataagcaactgctgcacctgggaaaaggaataatccaagccacat 1806
          |||||||
Db      1 CTAGACAGCAGATGCTGTCATTAAACAACTGCTCGGCCACCTGGAAAAAGGATAATCCAGGCACAT 60

Qy   1807 gggggacactgttgacccacacttcacgagtcgatgcacgacaacctcttctctgtggagtcaa 1866
          |||||||
Db     61 GGCGGACACTGTTGACCCCACTTCACGAGTCGATGGACGACAACCTCTTCTGTGGAGTCAA 120

Qy   1867 gtcaagaacgcgctatgtgcacaggcaataagagaaagaaagagtgttactgcacactgg 1926
          |||||||
Db    121 GTCAGCACGCCGCTATGCACAGGCAAATAAGAGAAAGAAAGAGCATGTGTTACTGCACACTGG 180

Qy   1927 ttaaaacacagtcttaaadaaaaaaatggtaccogcogcaccogagccccttcaactccca 1986
          |||||||
Db    181 TTAAAACACACTCTTAAAGAAAGAAGAAAAATGGTACCCCGCGCACCCGAGCCCTTCAACTCCCA 240

Qy   1987 gtggccttcccacagaggaagccatgttcacagacgatattattctgtgactggatt 2046
          |||||||
Db    241 GTGGCCTTCCCACCCAGGAAGAGGAAANGCCATGTTCACAGCATATTATTTCTGTGACTGGATTT 300

Qy   2047 gttgatagtgacagagatgacctaaaaataatggcttattttggcttattgccaaatatacg 2106
          |||||||
Db    301 CTTGATAGTGACAGAGATGACCTAAAAATTAATCGCTTATTTGGCAGTGTGCCAAAATATACC 360

Qy   2107 ggttatctatgcgcgcagcaacacagtcctcatcctgtaaagaaaaccaactggtttaaagtat 2166
          |||||||
Db    361 GGTTATCTATGCGCCGCACCAACACAGTCCTCATCTGTRAAAGAAAACCAACTGGTTTAAAGTAT 420

Qy   2167 gaaaagccaaagagtggaggaataccctgtgtcaacgcccagtggcttggccacatt ctt 2226
          |||||||
Db    421 GAANAAGCCAAAGAGTGGAGGAATACCCTGTGTCAACGCCCAGTGGCTTGGGCACATTCTT 480

Qy   2227 ctgggaaacttgaggcgactgaggcaagattccagttagtcgctacacgggcattcagtctg 2286
          |||||||
Db    481 CTCGGAAACTTTCAGGCACTCAGGCGACTTCAGTATAGTCGCTACACGGCCATTCAGTCTG 540

Qy   2287 caggalccattttgcccctaccaccagcgattagtttaaatctttagatgcttggaagtt 2346
          |||||||
Db    541 CAGGATCCATTTGCCCCTACCCCACCCACATTAGTTTTAAATCTTTAGATGCTTGGAAGTT 600

Qy   2347 cccctaaaagtgtctgcagaagtgttgatgagtaagactacctcccaaactgaaacag 2406
          |||||||
Db    601 CCCTTAAAAGTGTCTGCAGAAGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAG 660

Qy   2407 aatgaagtagtcaagtgtcccagccctctcccaaaagaggcacaatgaagagtaccacct 2466
          |||||||
Db    661 AATGAAGTAGTCAATGTCCCAGCCCTTCTCCCAAAAGAGGCACAATGAAGAGCTACCACCT 720

Qy   2467 cccactaaaaagctaactccagaattga-ccccttttgtgcttttcactgaattcgagcc 2525
          |||||||
Db    721 CCCACTAAAAAGCTAACTCCAGAATTGACACCCCCTTTTGTGTGCTTTTCACTGGATTCGAGCC 780

Qy   2526 tgtccaggttcacacagtatattaagagcatctctacatattcttg-gtgggagaggttggg-gag 2583
          |||||||
Db    781 TGTTCCAGGNNTCACACAGTATATTAAGAGCATCTCTANGNGTGGGAGAGGTTGCGNGAG 840

Qy   2584 tctgacac-agaagtgcacacactca-ttgccagcaaagt-gactcgaca-ccgtgaagtt 2639
          |||||||
Db    841 TCTGCACAAGAAGTGCACACACCTCANTTTTGCCAGCAAAGTGGACTCCCACCGTGAAGTT 900

Qy   2640 cctgacgcggatttctgtcgt 2660
          |||||||
Db    901 CCTGCACGGGGGATTTCTGTGGT 921
```

```
LOCUS         AI671887          760 bp    mRNA    linear    EST 17-DEC-1999
DEFINITION    wb41c03.x1 NCI_CGAP_GC6 Homo sapiens cDNA clone IMAGE:2308228 3'
              similar to TR:O15404 O15404 CAGF28 ;contains l:TR5.b2 l:TR5
              repetitive element ;, mRNA sequence.
ACCESSION     AI671887
VERSION       AI671887.1  GI:4851618
KEYWORDS      EST.
SOURCE        human.
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1  (bases 1 to 760)
  AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
  JOURNAL     Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
              R. Emmert-Buck, M.D., Ph.D.
              cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
              Bonaldo, Ph.D.
              cDNA Library Arrayed by: Greg Lennon, Ph.D.
              DNA Sequencing by: Washington University Genome Sequencing Center
              Clone distribution: NCI-CGAP clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              www-bio.llnl.gov/bbrp/image/image.html
              Insert Length: 1444   Std Error: 0.00
              Seq primer: -40UP from Gibco
              High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1..760
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2308228"
                     /clone_lib="NCI_CGAP_GC6"
                     /tissue_type="pooled germ cell tumors"
                     /lab_host="DH10B"
                     /note="vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not 1; Site_2: Eco R1; Plasmid DNA
                     from the normalized library NCI_CGAP_GC4 was prepared, and
                     ss circles were made in vitro. Following HAP purification,
                     this DNA was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from a pool
                     of 5,000 clones made from the same library (cloneIDs
                     1257096-1258631, 1469064-1470983, and 1475592-1476743).
                     Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      260 a  153 c  133 g  210 t  4 others
ORIGIN
```

```
Query Match    19.8%;  Score 707.2;  DB 9;  Length 760;
Best Local Similarity  96.3%;  Pred. No. 3.4e-107;
Matches 732;  Conservative  0;  Mismatches  27;  Indels  1;  Gaps  1;
```

```
Qy   2810 cactcttaaggcaaatatttttacatcaacacctgaatctgcccaagtcttccacta 2869
          |||||||
Db    759 CANTTTATTNAGGCAAAATATTTACATCACACTGGGATCCGCNCAAGTCNCTTACACTA 700

Qy   2870 tgaaggcaatcgtagagtgtgcaggagggaaagtgtgttatccaagcagcactcttcccgga 2929
          |||||||
Db    699 TGAAGGCCATCGTAGAGTGTCAGGGAGNAAAGTGTGTTATCCAAGCAGCCACTCTTCCCGNA 640

Qy   2930 agctcatggagcacaagcaagaactcgagtttgtccggaaataaatttaatatcctgtgaaa 2989
          |||||||
Db    639 AGCTCATGGAGCAC-AGCAGAACTCGAGTTTGTCGGAAATAAATTTAATATCCTGTGAAA 581

Qy   2990 atgacctcattatgccgagaaatatttttgccagagaggcatagatgttcacaatgcaagt 3049
          |||||||
Db    580 ATGACCTCATTTATGCCGAGAAATATTTTTGCCAGAGGCCATAGATGTTCACAATGCAGAGT 521

Qy   3050 tcgttctgactggagtgctcactcaaaacgctggacttatgaatcatatataagtttaactgat 3109
          |||||||
```

RESULT 3

LOCUS   AW073165   780 bp   mRNA   linear   EST 09-MAR-2000
DEFINITION wy94g07.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2556252 3',
            similar to TR:O15404 O15404 CAGγ28 ;contains LTR5.b2 LTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AW073165
VERSION   AW073165.1 GI:6028163
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 780)
  AUTHORS   NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute / National Institute of Neurological
            Disorders and Stroke, Brain Tumor Genome Anatomy Project
            (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 3034   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 455.
FEATURES   Location/Qualifiers
     source   1..780
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2556252"

Query Match   19.6%; Score 700.6; DB 9; Length 780;
Best Local Similarity   96.6%; Pred. No. 4.1e-106;
Matches 758; Conservative   0; Mismatches 21; Indels   6; Gaps 4;

Qy   2785 tccttaaaacgggcacacgttctccactcttaaggcgaaaatatttttacatacacacct 2844
           | |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   780 TCCCTTAAAAAGGGCACACGT---CCCNCCCTCNAAAGGCAAAATATTCTACATC-CACCT 725

Qy   2845 ggaatgtgcccaaagtctttccactatgaaggcaatcctgagatgtgtgcagggagagtg 2904
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   724 GGAATCTCCCCAAGTCTCCCACTATGAAGGCAATCGTAGAGTGTGCACCAGGAAGGTG 665

Qy   2905 tatccaagcagccatcttccggaagctcaatggagcacaagcaagaacccagagtttgtcg 2964
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   664 -TATCCAAGCAGCCATCTTCCGGAAGCTCAATGGAGCACAAGCAAGAACTCGAGTTGTCG 606

Qy   2965 gaaataaatttaatatcctgtgaaaatgaccttcattatgccgagaagaatatttgccaga 3024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   605 GAAATAAATTTAATATCCTGTGAAAAATGACCTTCATTTATCCCGAGAGAATATTTGCCAGA 546

Qy   3025 gccataggtgttcacatgcagagttcgttcactggagtgctcactcaaacgctgaac 3084
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   545 GGCATAGAATGTTCACAATGCAGAGTTCGTTCGACTGGAGTGCTCAACCGCTGCAC 486

Qy   3085 tatgaatcaatataagcttaactgatggcgtctagggctgcagtgaatgtcgaactcctggg 3144
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   485 TATGAATCAATATAAGTTAACTGATGGCGTCTCRAGGCTGCGCCGGTGCANTGCGAGACTCCTGCGG 426

Qy   3145 tgcgggcgtcggctctggcctcttcactgtgcagctccttcaccaatgctgctcttgtt 3204
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   425 TGCGGGCGCTGGCTCTTCCTGCCTGGCCGAGGAGCTCGTGCGCCGGTTCCTTCACATGCTGTGTT 366

Qy   3205 tccagtcgcttcctgggggatcagacgtgtgaagcaggagcagatatataataataatac 3264
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   365 TCCAGTCGCTTCCTTCCTGGGGGCATCAGACTGTGAAGCAGGAGCAGACAGATATAATAAATATAC 306

Qy   3265 tgcatctttttaagatgtgcaatttattctgaggaaacataaattatgtttggtattat 3324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   305 TGCATCTTTTTAAGATGTGCAATTTATTCTGAGGAAAACATAAATTAATGTTTGTATTAT 246

Qy   3325 atgactttaagg-cccacattaggtttttattgattcatttgccaggtttttaaatgtttt 3383
           |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   245 ATGACTTTAAGGAGCCCCCCATTAGGTTTTATTGANTTCATTTGCCCAGGTTTTTAAATGTTTT 186

Qy   3384 cacacaactgttacggagacttcaactagaaataaaatggtgtaaataagagacctgctat 3443
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 CACAAAACTGTTACGGGAGACTTCAACTAGAAATAAAAATGGTGCTAAAATAAAGCCCTTGCTAT 126

Qy   3444 ctctaaattattgatgttaaagattttgtgtactttgattgatttattcatcctt 3503
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   125 CTCTAAATTATGGATGTTAAAGATTTGAAAGTTTTGTGTACTTTGATTGATTTTTATTCTT 66

Qy   3504 atacttcgttttctttattgatatcttgcccacacattttaaaataaatgtactttttgaa 3563
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65 ATACTCTCGTTTTCTTTTATATTGATATCTTTGCCCACACATTTTAAAAATGTACTTTTTGAA 6

Qy   3564 cttag 3568
           |||||
Db   6

BASE COUNT   254 a   157 c   154 g   213 t   2 others
ORIGIN

Qy   40 TCTTGCCCACATTTTAAATAAATGTACTTTTTGAACTTAAA 1
           |||||||||||||||||||||||||||||||||||||||||
Db   3530 tcttgcccacatttaaataaatgtacttttgaacttaga 3569

Qy   100 CAAATCTTTTCTTACTTTGATTAT'TTTTATATACTCTCTTTTTTCTTTTATATTGATA 41
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3470 gaaatgttttgtacttgatttatttatctctttatctgttttcttttatattgata 3529

Qy   160 AGAAATAAAAATGGTGTAAATAAAGACCTTGCTATCTCTAAAATTATGGAATGTT 101
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3410 agaataaaaatggtgtaaataaagacctgctatctctaaattatgatgttaaagatt 3469

Qy   220 TTTTAPCATTTCAATTTCCAGGTTTTTAAANGTTTTTCACAAAACTGTTACGGGACTTCAACT 161
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3350 ttttatgatlcattgccaagtgtttcaccaaaactgttacgggacttcaact 3409

Qy   280 TAPTTCTCACCCAAACATAAATTAGTTTTGTCATTATATGACTTTAAGAGCCCACATTAGGT 221
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3290 tattctgaggaaacataaattatgtttggtattatatgactttaagagcccacattagt 3349

Qy   340 ACTGTGAAGCAGGAGAACAGGATATATAATAATAATACTGCATCTTTTTAAGATGTGCAATTT 281
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3230 actgtgaagcaggagacagagatataataataataactgcatctttttaagatgtgcaatt 3289

Qy   400 AGGAGCTCCTCCCCGTTCCTTCACCATCGTGTTTCCAGGTCTGCTTCCTGGGGGATCAG 341
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3170 aggagctgctcgtcctccatcacatgcttcctgtttccagctgctgttcctgggggatcag 3229

Qy   460 GGCGTCTAGCCTCGTCGCATCGCCAGCTCCTCCGGCGCGCCCGCTGCCGTGCTCGGCCTGGCC 401
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3110 ggcgctaggctccgctgctgctgactccggggggcgtgctgctgtctggctggctggcg 3169

Qy   520 TCGTTCTGACTGAGTCGTCACTCAAAACGCTGGACTGGAGTCAATCATATAAGGTTAACTGAT 461
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3050

```
Db     5 CTTAG 1
       |||||
```

```
RESULT   4
LOCUS    AI357401/c
LOCUS    AI357401                 811 bp    mRNA    linear    EST 15-FEB-1999
DEFINITION qy13d06.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2011883 3'
         similar to TR:O15404 O15404 CAGF28 ;, mRNA sequence.
ACCESSION AI357401
VERSION  AI357401.1  GI:4109022
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 811)
AUTHORS  NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE    National Cancer Institute / National Institute Of Neurological
         Disorders and Stroke, Brain Tumor Genome Anatomy Project
         (CGAP/BTGAP), Tumor Gene Index
JOURNAL  Unpublished (1998)
COMMENT  Contact: Robert Strausberg, Ph.D.
         Email: cgapbs-r@mail.nih.gov
         Tissue Procurement: David M. Louis, M.D., Myrna R. Rosenfeld M.D.,
         Ph.D.
         cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
         Bonaldo, Ph.D.
         cDNA Library Arrayed by: Greg Lennon, Ph.D.
         DNA Sequencing by: Washington University Genome Sequencing Center
         Clone distribution: NCI-CGAP clone distribution information can be
         found through the I.M.A.G.E. Consortium/LLNL at:
         www-bio.llnl.gov/bbrp/image/image.html
         Insert Length: 1784    Std Error: 0.00
         Seq primer: -40UP from Gibco
         High quality sequence stop: 364.
FEATURES             Location/Qualifiers
     source          1..811
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2011883"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGGGCGGGCCGCAATACTTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT      271 a    163 c    147 g    230 t
ORIGIN
```

```
Query Match  19.5%;  Score 698.2;  DB 9;  Length 811;
Best Local Similarity  94.2%;  Pred. No. 1e-105;
Matches 768;  Conservative  0;  Mismatches 43;  Indels  4;  Gaps  4;
```

```
Qy  2763 ttctcttcagcttggaagaatccttaaacacggcacacgtttctccactttaaggc 2822
         |||||| |||||||||||||| |||||||||| |||||||| |||||||||||||||  !
Db  811  TTTTTTTTTCAGCTCGAGAAAAATCCTAAAACGCGGCAC-GGTTCTTCCCCTCTTAAGCC 753
```

```
Qy  2823 aaaatatttttacatcaacaccctggaatcctgcccaaagtctttccactatgaaggcaatcgt 2882
         |||||| ||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  752  AAAATAATTTT-CATCACACTGGGAAAATCTGCCCAAGTCTTCCACTATAAGGCCCATTCGT 694
```

```
Qy  2883 agagtgtgcaggagggaaagtgttatccaagcagcccatcttccggaagctcatggaagca 2942
         |||||| ||| |||||||| ||||||||||||||||||||||||||||||||||||||||
Db  693  A-AGGGCGCAGGAGGAAAAGGTGTTATCAAAG-ACCCTTCTTCCGGAAGCTCATGGAGCA 636
```

```
Qy  2943 caagcagaactcgagttttgtcgggaaatatttaatatcctgtgaaaatgaacttcatt 3002
         |||||| |||| |||||||||||||||| ||||||||||||||||||||||||| |||||
Db  635  CAAGCAGAACTCGAGTTTGTCGGAAATATTTAATAATCCTGTGAAAATGAACTTCATTT 576
```

```
Qy  3003 atgccggagatatttgccagagaagtgtttcacaatgcagagttcagagttcgttctgactg 3062
         |||||| |||||||||||| ||||| |||||||||||||||||| |||||||||||||||
Db  575  ATGCCGGAGATATTTGCCAGAGAAGTGCCATGAAGTTCACAATGCAGAGTTCGTTCTGACTGG 516
```

```
Qy  3063 agtgctcactcaaacgctggactatgaatcatatataagtttaactgatggcgtctaggctg 3122
         |||||| ||||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db  515  AGTGCTCACTCAAACGCTGGACTATGAATCATATATAAGTTTAACTGATGGGCTCTAGGCTG 456
```

```
Qy  3123 ccgtgcatgtcgactcctgcgggtgcagggcgtggctgtctcttggggtcggaagagagctgctcg 3182
         |||||| |||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db  455  CCGTGCATGTCGACTCCTGCGGGTGCGGGCGTGGCTGTCTCTTGGGTCGGAGGAGAGCTGCTCG 396
```

```
Qy  3183 cttcctcaatgctcttgttcttccagctgctttctcggggggatcagactgtgaagcagg 3242
         |||||| ||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db  395  CTTCTTCACATGCTCTTGTTTTTCCAGCTGCTTTCCTCGGGGGATCAGACTGTGAAGCAGG 336
```

```
Qy  3243 aagacagatataataataatactgcatctcttttagagtgtgcaatttttattctgaggaa 3302
         |||||| |||| |||||||||||||||||||| ||||||||||||||||||||||||||||
Db  335  AAGACAGATATAATAATAATACTGCATCTCTTTTAGAGTGTGCAATTTTTATTCTGAGGAAA 276
```

```
Qy  3303 cataaaattatgttttgtattatatgagctccccaataggttttatgattcatt 3362
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  275  CATAAAATTATGTTTTGTTATATTATGAGCTCCCCCAATAGGTTTTATGATTCATT 216
```

```
Qy  3363 tgccaggttcttttaaatgtttcacacaaactgttacgggagctcaactagaaaataaaatgg 3422
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  215  TGCCAGGTTTTTAAATGTTTCACAAAACTGTTACGGGACTTCAACTAGAAAATAAAATGG 156
```

```
Qy  3423 tgtaaataaagaccttgctatctctcaaattatggatgttgaaagatttgaaaatgtttttgta 3482
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  155  TGTAAATAAAGACCTTGCTATCTCTAAATTATGGATGTTGAAAGATTTGAAAGTTTTGTA 96
```

```
Qy  3483 ctttgattatttttattcctgttttctttcttttattgatatcttgcccacatt 3542
         |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  95   CTTTGATTTATTTTTATTCCTGTTTTTCTTTCTTTTATTGATATCTTGCCCACATT 36
```

```
Qy  3543 ttaaaataaatgtacttttgaacttagaaaaaaaaa 3577
         ||||||||||||||||||||||||||||||||||||
Db  35   TTAAAATAAATGTACTTTTGAACTTAGAAAAAAAAA 1
```

```
RESULT   5
LOCUS    BE260052
DEFINITION 601150905F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503689 5',
         mRNA sequence.
ACCESSION BE260052
VERSION  BE260052.1  GI:9131032
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1  (bases 1 to 759)
AUTHORS  NIH-MGC http://mgc.nci.nih.gov/.
TITLE    National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL  Unpublished (1999)
COMMENT  Contact: Robert Strausberg, Ph.D.
         Email: cgapbs-r@mail.nih.gov
         Tissue Procurement: ATCC
         cDNA Library Preparation: Ling Hong/Rubin Laboratory
         cDNA Library Arrayed by: The I.M.A.G.E. Consortium
         DNA Sequencing by: Incyte Genomics, Inc.
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
         Plate: 1LCM177 row: i column: 02
         High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1..759
```

```
Qy  3063 ... LOCUS AI357401 811 bp mRNA linear EST 26-OCT-2000 759 bp mRNA linear
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503689"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT   212 a   179 c   177 g   191 t
ORIGIN

Query Match   18.2%;
Best Local Similarity 96.0%; Score 653.2; DB 10; Length 759;
Matches 724; Conservative 0; Pred. No. 2.8e-98;
                              Mismatches 23; Indels 7; Gaps 5;

Qy 1961 cgccgcaccgagcccttcacttcccagtagccttcccaccagagaggaaagccatgttcac 2020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 CGCCGCACCACCAGCCCTTCACTTCCCACGTAGCCTTCCCACCAGAGAGGAAAGCCATGTTCAC 68

Qy 2021 agcatattattctgtgactggatttgtgatgagtgacaagagatgacctaaaattaatgg 2080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 AGCATATTATTCTGTGACTGGA-TTGTTGATGATGATGACAAGAGATGACCTAAAATTAATGG 127

Qy 2081 cttattttggcagttgccaaaatacagggttatctatgccggagcaacacagtcctcatct 2140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128 CTTATTTGGCAGGTGCCAAAATATACGGGTTATCTATGCCGGAGCAACACAGTCCTCATCT 187

Qy 2141 gtaagacaacaactggttaaagtatgaaaaagccaaagaagtggaaggatacccgtgtca 2200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   188 GTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGATACCCTGTGTCA 247

Qy 2201 acgccaagtggctgggcgacattctctctgggaaacttgggcagtcagattcagt 2260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248 ACGCCCAGTGGCTTGGCGACAATTCTTCTGGGAAAC-TTGAGCACTGAGGCAGATTCAGT 306

Qy 2261 atagtcgctacacggccattcagtctcgaggatccattgccccctaccccagcacttagttt 2320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307 ATAGTCGCTACACGGCCATTCAGTCTCGCAGGATCCCATTGCCCCTACCCCAGCA-TTAGTTT 366

Qy 2321 taaatctttagatgcttggagagttcccttaaaagtgtctgcagagttgttgataagta 2380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367 TAAATCTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTA 426

Qy 2381 taagactacactcccaaactgaaacagaatgaagtagctaatgtccagcctcttcccaaa 2440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   427 TAAGACTACACTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCAGCCTCTTCCAAA 486

Qy 2441 gaccagaagaacgtaagacgtaccacctcccactaaaaagctaactccagaatgaccct 2500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   487 GACCCAGAAGTTGAAGACGTACCACCTCCCACTAAAAAAGCTAACTCCAGAATGGACCCC-T 545

Qy 2501 ttgtgcttcactggattcgagcctgtccacagttcaacagtatattaagaagctctaca 2560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   546 TTGTGCTTTTCACTCGGAATTCCAGCCTGTCCAGGTTCAACAGTATATTAAGAAGCTCTACA 605

Qy 2561 ttcttggtggagagggtgcggagtctgcacagaaagtgcacacacaactcattgccagcaaag 2620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   606 TTCTTGGTGGAGAGGGTTGCGGAGTCTGTGCACAGAAAGTGCACAGACACTTGCCAGCAAAG 665

Qy 2621 tgactcg-cacctgaagttcctgacggcgatt---tctgtcgtgaagcaacatagtgacg 2676
         |||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db   666 TGACTCGCCACCGGAAGTTCCTGACGGCGATTTCTGTCGTTGAAGCACATAGTGACG 725

Qy 2677 ccagagtggctggaagaatgcttcaggtgtcaga 2710
         | ||||||||||||||||||||| |||||| |||
Db   726 CCGAGTGGCCTGGAAAGAAATGGCCTCGGCTGTCACA 759

LOCUS   AI133662/c
LOCUS   AI133662   652 bp   mRNA   linear   EST 11-NOV-1999
DEFINITION HA2284 Human fetal liver cDNA library Homo sapiens cDNA, mRNA
            sequence.
ACCESSION   AI133662
VERSION   AI133662.1 GI:6360978
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 652)
AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Li,W., Wu,J., Zhou,S., Liu,M.
            and He,F.
TITLE   Expression profile analysis of a human fetal liver cDNA library
JOURNAL   Unpublished (1998)
COMMENT   Contact: Tong cao Tu
            Department of Hematology
            Beijing Institute of Radiation Medicine
            27 Taiping Road, Beijing 100850, P.R.China
            Tel: 0086-10-6815479
            Fax: 0086-10-6821 4653
            Email: yyt48@yahoo.com.
FEATURES   Location/Qualifiers
     source   1..652
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="Human fetal liver cDNA library"
                     /tissue_type="liver"
                     /dev_stage="fetal"
                     /lab_host="MC1061/P3"
                     /note="Vector: pCDNA1"
BASE COUNT   175 a   166 c   147 g   164 t
ORIGIN

Query Match   18.2%;
Best Local Similarity 99.8%; Score 650.4; DB 9; Length 652;
Matches 651; Conservative 0; Pred. No. 8.5e-98;
                              Mismatches 1; Indels 0; Gaps 0;

Qy 2558 acattcttggttggagaggttgcggagtctgcacagaaagtgcacacacctcattgccagca 2617
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   652 ACATTCTTGGTGGGAGAGGTTGCGGAGTCTGCACAGAAAGTGCACACACCTCATTGCCAGCA 593

Qy 2618 aagtgactcgcacctgaagttcctgacgcggcgatttctgtcgtgaagcaacatagtgacgc 2677
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   592 AAGTGACTCGCACCTGAAGTTCCTGACGCGGCGATTTCTGTCGTGAAGCACATAGTGACGC 533

Qy 2678 caaagtggctggaagaagtgcttcaggtgtcagaaagtccattgatgagcgagaactacatto 2737
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   532 CAGAGTGGCCTGGAACAAATGCTTCAGGTGTCAGAAGTTCATTCAATGAGCAGAACTACATTC 473

Qy 2738 tccgagatgctgagcgagaagtacttttctcctttcagcgttgaaagaatccttaaaacggg 2797
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   472 TCCGAGATGCTGAGCGAGAAGTACTTTTCTCCTTTCAGCGTTGAAAGAATCCTTAAAACGGG 413

Qy 2798 cacacgttctccactctttaaggcaaaatattttacatcacacctggaatctgcccaa 2857
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   412 CACACGTTTCTCCACTCTTTAACGCAAAAATATTTTACATCACACCTGGAATCTGCCCAA 353

Qy 2858 gtcttccacttgaagccaatcgtagagtgtgcagaggaggaaggtgttatccaagcagc 2917
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   352 GTCTTTCCACTTGAAGCCAATCGTAGAGTGGCCAGAGGAGGAAAGGTGTTATCCAAGCAGC 293

Qy 2918 catcttcccgaaagctcatgggcacaagcagaacctgagtgtttcggaaaataaatttaa 2977
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   292 CATCTTCCCGAAAGCTCATGGGCACAAGCAGAGAACCTGCAGTTTGTTCGGAAAATAATTTTAA 233

Qy 2978 tatcctgtgaaaatgacctcattatgccgagaatatttgcgaggaggcatagatgttc 3037
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   232 TATCCTGTGAAAATGACCTCATTATGCCGAGAATATTTGTCGGAAAATAATTTAA 3037

RESULT 7

LOCUS   AI741935   680 bp   mRNA   linear   EST 19-DEC-1999
DEFINITION wg29g05.x1 Soares_NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
            IMAGE:2366552 3' similar to TR:O15404 O15404 CAGF28 ;. mRNA
            sequence.
ACCESSION   AI741935
VERSION   AI741935.1 GI:5110143
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 680)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1113   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 432.
FEATURES   Location/Qualifiers
     source   1..680
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2366552"
                     /clone_lib="Soares_NSF_F8_9W_OT_PA_P_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from five normalized
                     libraries were mixed, and ss circles were made in vitro.
                     Following HAP purification, this DNA was used as tracer in
                     a subtractive hybridization reaction. The driver was
                     PCR-amplified cDNAs from pools of 5,000 clones made from
                     the same 5 libraries. The pools consisted of the following
                     libraries and cloneIDs: Soares NhHSF pool 1:
                     309384-310919, 323208-325895 Soares Nb2HP pool 1:
                     145032-147335, 147720-148103, 148872-149255, 15002 -
                     150407, 151176-152327 Soares Nb2HF8-9W pool 1:
                     758280-760583, 772104-774407 Soares NbHPA pool 1:
                     304776-306311, 320136-322823, 326280-326663 Soares NbHOT
                     pool 1: 723720-726407, 739080-740999 Subtraction by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT   225 a   138 c   116 g   201 t
ORIGIN

Query Match   17.9%; Score 639.8; DB 9; Length 680;
Best Local Similarity   96.7%; Pred. No. 4.7e-96;
Matches   653; Conservative   0; Mismatches 22; Indels   0; Gaps   0;

Qy   2906 tatccaagcagccatcttccgaagtcatggagcacaagcaagaactcgagtttgtcgg 2965
           || || | |||||||||||||||| || |||||||||| ||||||| |||||||||
Db   232 TATCCTGTGAAAATGACCTTCATTATGCCGAGAGGAGCACAGAGACTGGAGTTC 173

Qy   3038 acaatgcagagagttcgttcgactggagtgctcactcaaaagctggataataatata 3097
           | ||||| ||| ||| ||| ||||||| ||| || ||||||||||| || || |
Db   172 ACAATGCAGAGTTCGTTCTCGACTGGAGTGCTCACTCAAACCCTGGACTATGAATCATATA 113

Qy   3098 aagtttaactcgatggcgtctaggcgtgcccgtgcatcctgcggtgcgggggctggct 3157
           |||||||| ||||| |||||||||||||||| ||||| |||||||||||||| ||| ||
Db   112 AGTTTAACTCGATGGCGTCTAGGCGTGCCCGTGCATGCCTGCACTCCTCGCCATGCATGCGGGGCT 53

Qy   3158 gtctggctgcgtgcgggaggaggaggctgctgcgctcctccacatgctccttcctcttgttttccag 3209
           || ||| || |||| ||||||||||||||||||||||||||||||||||||||||
Db   52 GTCTGGCTGCGGAGGAGGAGGCTGCTGCGCTCCTCCACATGCTCCTTCCTCTTGTTTTCCAG 1

RESULT 8

LOCUS   BG900296   669 bp   mRNA   linear   EST 06-NOV-2001
DEFINITION HOA37-1-B1 HOA (Human Osteoarthritic Cartilage) Homo sapiens cDNA,
            mRNA sequence.
ACCESSION   BG900296
VERSION   BG900296.1 GI:14310545
KEYWORDS   EST.
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 669)
  AUTHORS   Kumar,S., Connor,J.R., Dodds,R.A., Halsey,W., Van Horn,M., Mao,J.,
            Sathe,G., Mui,P., Agarwal,P., Badger,A.M., Lee,J.C., Gowen,M. and
            Lark,M.W.
  TITLE   Identification and initial characterization of 5000 expressed
            sequenced tags (ESTs) each from adult human normal and
            osteoarthritic cartilage cDNA libraries
  JOURNAL   Osteoarthritis Cartilage 9 (7), 641-653 (2001)
  MEDLINE   21482651
COMMENT   Contact: Sanjay Kumar
            UW2109
            GlaxoSmithKline
            709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406, USA
            Tel: 610-270-7245

Qy   679 TATCCAAGCAGCCATCTTACGAGGAGGAGCTCATGGAGCACAAGAGAAACTCGAGTTTGTCGG 620
           ||||| ||| |||||||| ||||||| |||| ||||| ||| || ||| || ||||||||||
Db   2966 aaataatttataatcctgtgaaatgacctccattatgccgagaggagctcatgggcaac 3025

Qy   619 AAGTTAAYTCTAATATCCTGTAGAAATGACCTTCATGTATGCGAGAGTATTGTCGCAGAG 560
           ||| ||| || || | | | || || ||||||| ||| ||| |||||||||||
Db   3026 ggtagatgttcacaatgcagagttcgttctgactgaagtgctcactcaaacgctggact 3085

Qy   559 GCATAGAYGTTCACAATGCAGAGTTCGTTCGACTGGAGTGCTCACTCAAAGCGCTGGACT 500
           ||||| | |||||||||||||||||||||| ||||| ||||||||||| |||||||||||
Db   3086 atgaatcatataagtttaactgatgggcgtctaggcgtgcccgtgcatcctgcggt 3145

Qy   499 ATGAATCATATAAGTTTAACTGAGTTGGCGTCTAGGCGTGCCCGTGCATGTCGACTCCTGCGGT 440
           |||||||||||| |||||||| || ||| ||||||||||||||||||||| |||| |||||||
Db   3146 gcggggctggctgctggctgcgtgcgggaggaggaggctgctgcgctcctccacatgctccttctt 3205

Qy   439 GCGGGGCTGGCTGCTGGCTGCGTGCGGGAGGAGGAGGCTGCTGCGCTCCTCCACATGCTCTTGTTTT 380
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||| ||
Db   3206 ccagctgcttcctggggtagatcagactgtgaagcacaggagagacagtataaaataaatatatact 3265

Qy   379 CCAGCTGCTTCCTCGGGGATCAGACTGTGAAGCACAGGAGAGACAGATATAATAAAAATATACT 320
           ||||||||||||| ||||||||||||||||||||||| |||||||| || ||||| |||| |||
Db   3266 gcatctttttaagatgtgcaatttttattctgagaaaacataaattatgtgtttgtattata 3325

Qy   319 GCATCTTTTTAAGATGTGCAATTTTTATTCTGAGGAAACATAAATTATGTGTTTGTATTATA 260
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   3326 tgactttaagagcccacacattaggtttatgattcattgccaggtttttaaatgtttca 3385

Qy   259 TGGCTTTAAGAGCCCACACATTAGGTTTATGATTCAYTGGCCAGGTTTTTAAATGTTTTCA 200
           || ||||||||||||||||||||||||||||||||| || ||||||||||||||| |||||
Db   3386 caaaactgttacgggactcaactagaaataaaatgtgtaaatataagacctgctatct 3445

Qy   199 CAAAACTGTTACGGGACTTCAACTAGAAATTAAAATGGGTGTAAATAAAGACCTTGCTAYCT 140
           |||||||||||||||| ||||||||||||| |||||| |||||||| ||||| ||||| |||
Db   3446 ctaaatattgatgttaaagattgaaatgtctttgtacttgactcattctttatttcttat 3505

Qy   139 CTAAATATTGAYGTTAAAGATTGAAATGTGTACTTGACTTTGAATTTTTTATTTCTTAT 80
           |||||||||| |||||||||||||||||| | ||||||| ||| ||||||||||||||
Db   3506 acctgtgttttcttttttatgatatcttgcccacatttaaaataaatgtacttttgaact 3565

Qy   79 ACYTGTGTTTCTTTTTATTATGATATCTTGCCCCACATTTAAAATGAATGTACTAAAGAACT 20
           || |||||||||||||||| ||||||||||||| ||||||||||| || ||||| |||||||
Db   3566 tagaaaaaaaaaaa 3580

Qy   19 TAAAAAAAAAAAAA 5
           ||||||||||||
Db   19 TAAAAAAAAAAAA 5

Left column (RESULT 9)

```
                  Fax: 610-270-5598
                  Email: sanjay_kumar-1@gsk.com
                  Seq primer: T7.

FEATURES          Location/Qualifiers
     source       1..669
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone_lib="HOA (Human Osteoarthritic Cartilage)"
                  /tissue_type="cartilage"
                  /lab_host="E.coli DH10 B"
                  /note="Vector: pSPORT 1; Site_1: SalI; Site_2: NotI;
                  Directional"

BASE COUNT      200 a    179 c    159 g    130 t     1 others
ORIGIN

Query Match    16.3%;    Score 583;    DB 10;    Length 669;
Best Local Similarity    99.8%;    Pred. No. 1.2e-86;
Matches 583; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1003 ggggaattaatgttgatgatcttcaagattcatccaccggaaaaacaggaggagaaatta 1062
        ||||||||||| ||||||||||||||||||||||||||| |||||||||| ||||||| ||
Db   86 CGGGAATTAATGTTTGATCGATTTCATCACCGGAAAAACAGGAGAGAAATTTA 145

Qy 1063 aacttgaaccccggcccgaagtcccacagttagctgcagcaaagtccaggga 1122
        ||||| |||||||||||||||| || ||||||||||||||||||||||||||
Db  146 AACTTGACCCCCGGCCCGAAGTCCCACAGTTAGCTGCAGCAAAGTCCAGGGA 205

Qy 1123 aaggagacctgggttgattaacttgtgtgccaatgtcccagtgtaacatttg 1182
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  206 AAGGAGACCTGGGTTGATTAACTTGTGTGCCAATGTCCCAGTGTAACATTTG 265

Qy 1183 cccctgagggtccgggttaattttaatggtgtgttggacaaacctccaaagttcctgaaaga 1242
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  266 CCCCTGAGGGTCCCGGGTAATTTAATCGCGTGCTGGACAAAAACCTCCAAAGTTCTGAAAGA 325

Qy 1243 tcaggaaatgatgagctacctggagtccagctgtacggacactgaggaattcttactaataat 1302
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  326 TCAGAAATGATGAGCTACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAAT 385

Qy 1303 gctgacattcagcagagtgaaccggcccatccaaatgtagacacatatcttcacgagactcttta 1362
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  386 GCTGACATTCAGCAGAGTGAACCGGCCCATCAAATGTAGACACATATCTTCACGAGACTCTTTCA 445

Qy 1363 gcacctacgaaaaatttagaacacaggtgaatacaagccagcaggcacacaaatgcc 1422
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  446 GCACCTACGAAAAATTTAGAACAGAGGTGAATACAAGTGCAGCAGGACAATACAAATGCC 505

Qy 1423 aatgacagtgctgttttagccaagtgaaagtgactccagagacacacatgctacagcagcag 1482
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  506 AATGCACGTGCTGTTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCAGCAG 565

Qy 1483 cagcaggcccagcagcagcagcagcagcagcccgtttacacctttacacctttcagcccagcagata 1542
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  566 CAGCAGGCCCAGCAGCAGCAGCNGCAGCGACACCCGGTTTACACCTTTCAGCCCCAGCAGATA 625

Qy 1543 atgcagctccagcagcagcagcagcagcagagatctctcagcaac 1586
        ||||||||||||||||||||||||||||||||||||||||||
Db  626 ATGCAGCTNCAGCAGCAGCACCACCAGCAGCAGAATCTCTCAGCAACC 669
```

Right column

```
RESULT    9
LOCUS        AU123803        625 bp   mRNA    linear    EST 23-OCT-2000
DEFINITION   AU123803 NT2RM2 Homo sapiens cDNA clone NT2RM2001075 5', mRNA
             sequence.
ACCESSION    AU123803
VERSION      AU123803.1   GI:10948519
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 625)
  AUTHORS    Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
             Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
             Isogai,T.
  TITLE      HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
             Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki
             ,Y., Sugano,S., Isogai,T.)
  JOURNAL    Unpublished (2000)
COMMENT      Contact: Takao Isogai
             Helix Research Institute
             Genomics Laboratory
             1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
             Tel: 81-438-52-3951
             Fax: 81-438-52-3952
             Email: genomics@hri.co.jp
             HRI human cDNA project; 5'- & 3'-end One pass sequencing; Helix
             Research Institute; cDNA library construction: Department of
             Virology, Institute of Medical Science, University of Tokyo, and
             Helix Research Institute.

FEATURES          Location/Qualifiers
     source       1..625
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="NT2RM2001075"
                  /clone_lib="NT2RM2"
                  /cell_type="teratocarcinoma"
                  /cell_line="NT2"
                  /note="Vector: pME18SFl3; mRNA from uninduced NT2 neuronal
                  precursor cells"

BASE COUNT      198 a    154 c    141 g    126 t     6 others
ORIGIN

Query Match    16.2%;    Score 581.2;    DB 9;    Length 625;
Best Local Similarity    98.5%;    Pred. No. 2.3e-86;
Matches 605; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

Qy  941 aagaaaggtctcctcagtgaccagcagtttcacctaaatccaacactgaaaaatcta 1000
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2 AAGAAAGC.TCTCCTCAGTGACCAGCAGTTTTCACCTAAATCCAACACTGAAAAATCTA 61

Qy 1001 aagggaaattaatgtttgatgattcttcagagattcatcaccggaaaaacaggaggagaaatt 1060
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 AAGGGGAAATTAATGTTTGATGATTCTTCACCTAAATCATCACCGGAAAAACAGGAGGAGAAATT 121

Qy 1061 taaactggaccccggccgggaagtcccacagttagctgcagcaaaagcccaggctgcctcagg 1120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  122 TAAACTGGACCCCGGCCGGGAAGTCCCACAGTTAGCTGCAGCAAAACCGCAGGCTGCCTCAGG 181

Qy 1121 gaaaggaggacctgggttgattaacttgtgtgccaatgtcccacccgtcccagtaacattt 1180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 GAAAGGAGACCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTT 241

Qy 1181 tgcccccctgaggtccgggttaatttaatggctgctgggacaaaaacctccaaaagttctgaaa 1240
        ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  242 TGCCCCCTGAGGTCCGGGTAATTTAATGGCTGCTGGGACAAAAACCTGCAAAGTTCTGAAA 301

Qy 1241 gatcaagaaaatgata-gctacctggagtccagctgtacggacacctagactatctttacaagactctt 1299
        ||||||||||||||| |||||||||||||||||||||||||||| |||||||||||||||
Db  302 GATCAGAAAATGATAATGCTACCTGGAGTCCAGCTGTACGGACCATCAAATGTACAGACTCTT 361

Qy 1300 aatgctgacattcagcagagtgaaccggccccatcaaaatgtagcacatatctttacaagactctt 1359
        ||||||||||||||||||||||||||||||||||| ||||||||||| |||||||||||
Db  362 AATGCTGACATTCAGCAGATGAACCGGCCCATCAAAATGTAGCACATATCTTACAGACTCTT 421

Qy 1360 tcagcacctacgaaaaatttagaacagcaggtgaatacaagccagcaggacacaacaat 1419
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 TCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGNCACAGGGACATACAAAT 481

Qy 1420 gccaaatgc-agtgtcgtgtttagccaagtgaaagtgactccagagacacacatgctacagca 1478
        ||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  482 GCCAAATGCAAGTGCTGNTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACAGCA 541
```

```
Qy   1479 gcagcagcagcagcccagcagcagcagcagcaccggtttacacacttcagcccagca 1538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 GCAGCAGCAGCCCCAGCCAGCAGCAGCCACCACCCCGTTTTACACCTTCAGCCCCACCA 601

Qy   1539 gataatgcagctcc 1552
          |||||||||||||||
Db    602 GATAATGCANCTTC 615


RESULT 10
LOCUS        BE541477
DEFINITION   601067929F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:3454183 5',
             mRNA sequence.
ACCESSION    BE541477
VERSION      BE541477.1  GI:9770122
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 803)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: Incyte Genomics, Inc.
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM8438 row: j column: 08
             High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1..803
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3454183"
                     /clone_lib="NIH_MGC_10"
                     /cell_line="MGC36"
                     /lab_host="DH10B"
                     /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5 kb. Library prepared by Life
                     Technologies."
BASE COUNT    188 a    207 c    181 t    227 g
ORIGIN

Query Match     16.0%;  Score 573.2;  DB 10;  Length 803;
Best Local Similarity   90.1%;  Pred. No. 4.6e-85;
Matches 726;  Conservative  0;  Mismatches  68;  Indels  12;  Gaps  10;

Qy   2009 agccatgttcacagatcattattctgtgactgattcgttgatagagtgacaagatgacc 2068
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCCATGTTCACAGCATATATTATTTCTGTGACTGGA-TTGTTGATACTGACAGAGATACC 59

Qy   2069 taaaattaatgcttcattggcaggtgccaaatatacggttatctatccgcagcaaca 2128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     60 TAAAATTAATGCTTCACAGGTGCCAATATACGGTTATCTATCCGCAGCAAACA 119

Qy   2129 cagtcctcactgtaaagaaccaactcgsttaaagtatgaaaaagagagtggaggaa 2188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    120 CAGTCCTCACTGTAAAGAACCAACTGGTT-AAGTATGAAAAAGCCAAAGAGTGGAGGA 179

Qy   2189 taccc-tgtgtcaacgcccagtggccttggcgccaattcttggaaaatttgaggcactg 2247
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    180 TACCCGTGTTGTCAAGCGCCCA--CGGCTCGGCGGACAATTCTTCTCGGGAAACTTTGAGGCACTG 237

Qy   2248 aggcagattcagtatagtcgctacacaggcattcagttctgcaggatcctcattgccctctacc 2307
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    238 AGGCAGANGTCAGTATAGTCGGCTACACGGCGATTCAGTCTGCAGGGATCCATTGGCCCCTACC 297

Qy   2308 cagcacattagttttaaatcttttagatgtgttgagagttccctgaaaagtgtctgcagag 2367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298 CAGCACATTAGTTTNAAACCTTTTAGATGC-TGGAGAGTTNCCTTAAAAGTGTCTGCAGAG 356

Qy   2368 ttgttgatgagtatalagacttacctcccaaactga-aacagaagatgaagtagctaatgtcca 2426
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    357 TTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAGAAGATGAAGTAGCTAATGTCCA 416

Qy   2427 gccttcttcccaaagagcagaattgaagacgtacacccctcccactaaaaagctaactcc 2486
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    417 GCCTTCTTCCCAAAGAGCCAGAATGGAAGACGTACCCACCTCCCACTAAAAAGCTAACTCC 476

Qy   2487 agaaatgaccccttttgtgctttcactggattccagcctgtccacagttcacacagtatat 2546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    477 AGAAATGGACCCC-TTTGTGCTTTTCACTGGATTCCGACCCGTCCAGGTTCACACAGTATAT 535

Qy   2547 taagaagctctacacattccttggtcggagaagttgcggaggtctgcacagaagtgcacacac-c 2605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    536 TAAGAAGCTCTACAATTCTTGGTGGAGAGGTGCGGAGGTCCGGACGTCCTCGCACAGAAGTGCACACAC 595

Qy   2606 tcattgccagcaaagtgactcgcaccgtgaaagttcctgacggcgcattcttctgtgaagc 2665
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    596 CTCATCGGCAGCAAAGTGACTCGCACCGTGAAAGTTCCGGACGGGGCATTTC-GGCGTGAAGC 654

Qy   2666 acatagtgacgccaga--gtggctggagaatgcttcaggtgtcagaagttcattgatga 2723
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    655 ACATAGTGACGGCAGAAGGGGGCTGGAGAAGCCCCACGCGGGGCCGAAGTTCCTTGGGGGA 714

Qy   2724 gcagaactacattctccgagatgctgaggcagaagaagtactttctccttcaacttggaaga 2783
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    715 GCCGGAACTACAAATTCCCGAGAAAAGCGGAAAGCCCGAAG-GCAATTCTCCTCCAACGGGGCGAAAA 773

Qy   2784 atccttaaaacgggcacacgttttctc 2809
          |||||||||||||||||||||||||||
Db    774 ACCCTGAAAAGGGGCGCACGGGTTCCCC 799


RESULT 11
LOCUS        AW009159/c
DEFINITION   597 bp    mRNA    linear    EST 08-MAR-2000
             ws78b09.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2504057 3'
             similar to TR:O15404 O15404 CAGF28 ;contains LTR5,b2 LTR5
             repetitive element ;, mRNA sequence.
ACCESSION    AW009159
VERSION      AW009159.1  GI:5857937
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 597)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
             , Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Insert Length: 854   Std Error: 0.00
             Seq primer: -40UP from Gibco
             High quality sequence stop: 370.
FEATURES             Location/Qualifiers
     source          1..597
```

RESULT 12
LOCUS   AW004836/C
DEFINITION   ws93c05.xl NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2505512 3',
             similar to TR:015404 015404 CAGF28 ;contains LTR5.b2 LTR5 LTR5
             repetitive element ;, mRNA sequence.
ACCESSION   AW004836
VERSION   AW004836.1   GI:5855614
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 552)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
             Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.
             , Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             www-bio.llnl.gov/bbrp/image/image.html
             Seq primer: -40UP from Gibco
             High quality sequence stop: 390.
FEATURES   Location/Qualifiers
     source   1..552
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2505512"
                     /sex="pooled"
                     /tissue_type="colon"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from 12 pooled bulk tumor samples and primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library went through one round of
                     normalization."
BASE COUNT   201 a   112 c   94 g   145 t
ORIGIN

Query Match   15.3%; Score 547.2; DB 9; Length 552;
Best Local Similarity   99.5%; Pred. No. 9.9e-81;
Matches 549; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3016 tttgccagaggcaatagatgttcacaatgcaagagttcgttctgactgggagtgctcactcaa 3075
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   552 TTTGCCAGAGGCATAGAIGTTCACAAATGCAGAGTTCGTTCGGAGTGCTCACTCAA 493

Qy 3076 acgctggactatgaatcataagattaactgatggcgtctaggctgcctgcatgtcga 3135
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   492 ACCCTGGCA^TATGAATCATATAAGTTTAACTGAIGGCGTTCTAGGCTGCCTGCATGTCGA 433

Qy 3136 ctcctgcggtgcggggctggctctggctggccaggaggagctgcgcttcctcacatg 3195
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   432 CTCATGCGGTGCGGGGCTGGCTGTCTGGCTGGCCAGGAGGAGACCTGCGCCGCTTGCCTCACATG 373

Qy 3196 ctcttgttttccagctgcttcctggggggatcagaactgtagaggaaacagaatataa 3255
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   372 CTCTTGTTTTCCAGCTGCTTCCTGGGGGGATCAGAACTGTGAAGCAGGAGAGACAGAATATAA 313

Qy 3256 taaatatactgcatctttttaagatgtgcaatttttattcctgaggaaacataattattatgtt 3315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   312 TAAATATACTGCATCTTTTTAAGAGTGTGCAATTTTATTCCTGAGGAAACATTAAATTATTATGTT 253

Qy 3316 ttgtatatatgacttaagagcccacaattagagttttatgatcattgccaggtttta 3375
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   252 TTGTATTATATGACTTTTAAGAGCCCACAATTAGGTTTTTATGAITCANTTTGCCAGGTTTTTA 193

Qy 3376 aatgtttcacacaaactgttacgggagcttcaactagaaaataaaatggtgttaaataaagac 3435
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 AATGTTTCACAAAAACTGTTACGGGACTTCAACTGAAATAAAAIGGTGTAAAATAAAAGAC 133

Query Match   15.6%; Score 558.4; DB 9; Length 597;
Best Local Similarity   97.1%; Pred. No. 1.4e-82;
Matches 579; Conservative 0; Mismatches 16; Indels 1; Gaps 1;

Qy 2975 taatataccttgaaaatgaacctlcaittatgccgagagaatatttttgccagaggcaatagatg 3034
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   596 TAAGACCCTGTGAGCAGACACCCTTCCTTTAAGCCAAAAATAATTTTGCCACAGGCATAAATG 537

Qy 3035 ttcacaatgcaagagttcgttctgactggagtgctccactcaaacgtggactatgaatcat 3094
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   536 TTCACAAATGCAGAGNAGAATCGTTCTGACTGAAGTGCTCACTCAAACCGTGGACTAIGAATCAT 477

Qy 3095 ataa-gttaactgatgggtctaggctgcctgcacatgtcgactcctgcgggttgcggggct 3153
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   476 ATAAGCTTTAACTGAIGGCGTCTAGGCTGCCGTGCCGTCCATGTCGACTCCTGCGGGTGCGGGGCT 417

Qy 3154 ggctgctcggctgcggaggagctgcgtgtgcgccttcctcacatgctcttgttttccagctgc 3213
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   416 GGCTGCTGCTCGGCTGGCGAGGAGCTGCGTGCTGCCTTCCACATGCTCTTGTTTTCCAGCTGC 357

Qy 3214 ttcctgggggatcagaactgtgaagcgacaggaagaaccagatataataataataactactgcatctt 3273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   356 TTTCCTGGGGGATCAGACTGTGAAGCAGGAGAAGACAGATATAATAAATAAATACTGCATCTTT 297

Qy 3274 ttaagatgtgcaatttattctgaggaaaacataaatatgttttgtatttgtattatatgactta 3333
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   296 TTAAGATGTGCAATTTTATTCTGACGAAAACATAAAATTATTCTTGAGGAAACATGAICGACTTTA 237

Qy 3334 agagcccacattagcttgtttattaatgttttaaatgttttcacaaaactg 3393
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   236 AGAGCCCACATTAGGTTTTATTGATTCATTTTGCCAGGTTTTTCACAAAACTG 177

Qy 3394 ttacgggacttcaactagaaaataaaatggtgtaaataaagacccttgctatctctaaatta 3453
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176 TTACGGGACTTCACTGAAAATAAAAAAAGACCTTGCTAATCTCTGCTATCTCTAAATTA 117

Qy 3454 tggalgttaagatttgaaatgtttgtacttgattattttttattcttctctgtt 3513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   116 TGGATGTTAAAGATTTGAAATGTTTGTACTTTGCAITTATTTTATTTCTTATACTCTGTT 57

Qy 3514 tccttttattttgatatcttgcccacacattttaaataaatgttacttttgaacttaga 3569
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56 TTCTTTTATAITATTGATAATCTTGCCCACATTTAAAAIANTGTACTTTTGAACTTAAA 1

RESULT 13

LOCUS   BF939756   542 bp   mRNA   linear   EST 22-JAN-2001
DEFINITION   nac83a11.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3440924 3'
                similar to TR:O15404 O15404 CACF28 ;contains LTR5.b2 LTR5
                repetitive element ;, mRNA sequence.
ACCESSION   BF939756
VERSION   BF939756.1 GI:12356999
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 542)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute / National Institute of Neurological
                Disorders and Stroke, Brain Tumor Genome Anatomy Project
                (CGAP/BTGAP), Tumor Gene Index
  JOURNAL   Unpublished (1998)
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
                Ph.D.
                cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
                Bonaldo, Ph.D.
                cDNA Library Arrayed by: Greg Lennon, Ph.D.
                DNA Sequencing by: Washington University Genome Sequencing Center
                Clone distribution: NCI-CGAP clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL, send email to:
                info@image.llnl.gov
                Seq primer: -40UP from Gibco
                High quality sequence stop: 475.
FEATURES   Location/Qualifiers
     source   1..542
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3440924"
                     /clone_lib="NCI_CGAP_Brn23"
                     /tissue_type="glioblastoma (pooled)"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGAGACGGGCCGCATATCTTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M. Fatima Bonaldo."
BASE COUNT   198 a   109 c   90 g   145 t
ORIGIN

Qy   3027 catagaatgttcacaatgcacagagtcgttctgactggagtgctcactcaaacgctggacta 3086
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Matches 541; Conservative 0; Mismatches   1; Indels   0; Gaps   0;
Best Local Similarity   99.8%; Pred No. 1.3e-79;
Query Match   15.1%; Score 540.4; DB 10; Length 542;

Qy   3436 cttgctatctctaaaattatggatgtgttaaagatttgaaatgttgtactcttgattattt 3495
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 CTTGCTATCTCTAAATTATGGAGATTTAAAGATTTGAAATGTTTGTACTTGATTATTTT 73
Qy   3496 tatttcttatctctcgtttctctttttctctttatcatcttgcccacattttaaataaatgta 3555
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72 TATTTCTTTATCTCTCGTTTCTCTTTTTATTATTGATATCTTGCCCACATTTTAAATCCATGTA 13
Qy   3556 cttttgaacta 3567
          |||||||||||
Db   12 CTTTTGAACTTA 1

RESULT 14

LOCUS   AI807116   598 bp   mRNA   linear   EST 18-DEC-1999
DEFINITION   we85g11.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
                IMAGE:2347940 3' similar to TR:O15404 O15404 CACF28 ;contains
                LTR5.b2 LTR5 repetitive element ;, mRNA sequence.
ACCESSION   AI807116
VERSION   AI807116.1 GI:5393682
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 598)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                This clone is available royalty-free through LLNL ; contact the
                IMAGE Consortium (info@image.llnl.gov) for further information.
                Insert Length: 1375   Std Error: 0.00
                Seq primer: -40UP from Gibco
                High quality sequence stop: 408.
FEATURES   Location/Qualifiers
     source   1..598
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2347940"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Db   542 CATAGAATGTTCACAATGCACAGAGTTCGTTCTGACTGGAGTGCTGCTCACTCAAACGCTGGACTA 483
Qy   3087 tgaatcatataaagttaactgatgacgcgtctaggctgacctgcatgtgatctcgactcgcggtg 3146
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   482 TGAATCATATAAGTTAACTAGTGACGCGTCTAGGCTGCCATGTGCGATGTCGACTCGCGGTG 423
Qy   3147 cggggctggctgctgtcgtggcgagagagtgctgcgcttccttcacaatgctgctcttgtttc 3206
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   422 CGGGGCTGGCTGTCGTCCTGGCGAGGAGAGCTGCGGCTTCCTTCACAATGCTGCTCTTGTTTC 363
Qy   3207 cagctgctttcctgggggatcaagctgtgaagcaggaagcaggagaagaacagatataataaatactg 3266
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   362 CAGCTGCTTTCCTCGGGGATCACAGACTGTGAAGCACGAAGCAGCAGAGATATATAAAATACTG 303
Qy   3267 catcttcttttaaagatgtgcaatttttattctgagggaaacatatatgttttgtattatat 3326
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 CATCTTCTTTTAAGATGTGCAATTTTATTCTGAGGAAACATAAATTATGTTTTGTATTATAT 243
Qy   3327 gactttaaggagagcccacattagggttttttatgattcattttgccaggtttctttaaatgttttcac 3386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 GACTTTAAGAGAGCCCACATTAGGTTTTTATGATTCATTTTGCCAGGTTTTTAAATGTTTTCAC 183
Qy   3387 aaaactgttacggggacttcaactagaaatacagtggtgtaaatacaagaccttgctatctc 3446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   182 AAAACTGTTACGGGACTTCAACTAGAAATAAAATGGTGTAAATACAAGACCTTGCTATCTC 123
Qy   3447 taaattattggatgtaaagatttgaaatgtttgtactttgattattttattcttctata 3506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 TAAATTATAGGATGTAAAGATTTGAAATGTTTGTACTTTGATTATTTTATTTCTTTATA 63
Qy   3507 ctctgttcttcttttatatgatatctctgcccacatttttaaataaatgtacttttgaactt 3566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62 CTCTGTTTCTTTTTAATTGATAATACTTGCCCACATTTTAAATAAATGTACTTTTGAACTT 3
Qy   3567 ag 3568
          ||
Db   2 AG 1

BASE COUNT   215 a   119 c   105 g   159 t
ORIGIN

Soares and M. Fatima Bonaldo. "
726408-728711, and 729096-731399. Subtraction by Bento
I.M.A.G.E. clones 297480-302087, 682632-687239,
from the same 3 libraries. The pools consisted of
was PCR-amplified cDNAs from the pools of 5,000 clones made
tracer in a subtractive hybridization reaction. The driver
vitro. Following IIAP purification, this DNA was used as
NCI_CGAP_GCB1) were mixed, and ss circles were made in
libraries (fetal lung NbHL19W, testis NHT, and B-cell
Equal amounts of plasmid DNA from three normalized

Qy   2971 atttaattaatatcctgtgaaaatgacctteattatgccgagaatatattttgccaagaggcata 3030
           |||||| ||| |||||| |||||| ||||||| ||||| |||| | ||||||| ||||||||
Db   598 ATTTTAAATTCCTGTGAAAATGCCTTCATTTATCCGAAGAGAATATATTTTGCAGAGGGCATA 539

Qy   3031 gatgttcacaatgcagag-ttcgttctgactggagtgctcactcaaacgctggactatga 3089
           |||||| ||||||| ||| ||||||||||||||||||||||| |||||||||||||||||
Db   538 CATGTTCACAATCCACAGAGTTTCGTTCTGACTGGAGTGCTCACTCAAACCGCTGGACTATGA 479

Qy   3090 atcatalaagtttaactcgatggtgcatgtgccgtccatgtcgactcctgcgggtgcg 3149
           || | |||||| |||||||| ||||||| || |||||||||||| ||||||| |||||
Db   478 ATCATATAAGTTTAACTCATGGTGCGTCTACGGCGTCCCATGTCGACTCCTGCGGGTGCG 419

Qy   3150 ggctggctgtctggctggcggagggagctgctcactcacctgctcttgtttccag 3209
           |||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   418 GGCTGGCTGTCTGGCTGGCGGAGGAGCTGCTCCGCTTCCTTCACATCCTCTTGTTTCCAG 359

Qy   3210 ctgctttcctgggggatcagacgtgaagcagaagacagagatataataataactgcat 3269
           ||||||||| ||||||||| || |||||||||| ||||| |||||||||| |||||||
Db   358 CTGCTTCCTCGGGGATCAGACTGTGAAGCAGAAGCAGGAGATATAAATATACTGCAT 299

Qy   3270 ctttttaagatgtgcaatttlattctcaagagaaaaaataattatgttttgtatatatgac 3329
           |||||||||||||||||| ||||| ||||||||| |||||||||||| |||||||||||
Db   298 CTTTTTAAGATGTGCAATTTTATTCTGAGGAAAACAATAAATTAATGTTTTGTATTATATGAC 239

Qy   3330 tttaagagcccacattaggttttatgattcattgccaggtttttaaatgtttccacaaa 3389
           |||||| ||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   238 TTTAAGAGCCCACATTAGGTTTTATGATTCATTGCCAGGTTTTTAAAAGTTTTCACAAA 179

Qy   3390 actgtacggagctcaactagaaaataaatcgtgtaaataaagacctgctatcctctaa 3449
           |||||||||||||||||||||||||||| |||||||||||||||||||| |||||
Db   178 ACTGTTACGGCACTCAACTAGAAAATAAAATGTCGTAAATAAAGACCTTGCTATCTCTAA 119

Qy   3450 attatggatgtgtaaagagattgaaatgltttttactttaattgatttatttttattcttatatc 3509
           ||||| ||| |||||| |||||| ||||
Db   118 ATTATGGATGTTAAAGATTTGAAATGTTTTCTTACTTTGATTAATTTTATTTCTTTACACTC 59

Qy   3510 tgtttctcttttttatattgatatctgcccacacattttaaaataaataaatgtactttttgaactta 3567
           ||||| |||||||||||||||||| || || |
Db   58 TGTTTTCTTTTATATCCCTATCTTGCCCCATTTTTAAATAAAATGAACTTTGGACCTA 1

ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
SOURCE   Homo sapiens (human).
KEYWORDS   EST.
VERSION   AI393049.1 GI:4222596
ACCESSION   AI393049
            similar to TR:O15404 O15404 CAGF28 ;contains LIR5.b2 LIR5
            repetitive element ;; mRNA sequence.
DEFINITION tg25c09.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE.2109808 3',
LOCUS   AI393049   551 bp   mRNA   linear   EST 30-MAR-1999
AI393049/c

RESULT 15

Query Match   15.1%;   Score 539.6;   DB 9;   Length 598;
Best Local Similarity   95.0%;   Pred. No. 1.8e-79;
Matches 568; Conservative   0; Mismatches 29;   Indels 1;   Gaps 1;

(right column)

AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
           National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE   Tumor Gene Index
           Unpublished (1997)
JOURNAL   Contact: Robert Strausberg, Ph.D.
COMMENT   Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
           M.D., Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNl at:
           www-bio.llnl.gov/bbrp/image/image.html
           Insert Length: 738   Std Error: 0.00
           Seq primer: -40UP from Gibco
           High quality sequence stop: 359.

FEATURES   Location/Qualifiers
    source   1..551
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_lib="NCI_CGAP_CLL1"
              /clone="IMAGE.2109808"
              /tissue_type="B-cell, chronic lymphotic leukemia"
              /lab_host="DH10B"
              /note=Vector: pT7T3D-Pac (Pharmacia) with a modified
              polylinker; Site_1: Not I; Site_2: Eco RI; primer
              was primed with a Not I - oligo(dT) primer [5'
              TGTTACCAATCTGAAGTGGACGAGCGCGGCCGCTTTTTTTTTTTTTTTTTTT
              T 3']; double-stranded cDNA was ligated to Eco RI
              adaptors (Pharmacia), digested with Not I and cloned into
              the Not I and Eco RI sites of the modified pT7T3 vector.
              Library is normalized, and was constructed by Bento
              Soares and M.Fatima Bonaldo. "

BASE COUNT   202 a   108 c   103 g   138 t
ORIGIN

Query Match   14.8%;   Score 528.6;   DB 9;   Length 551;
Best Local Similarity   97.5%;   Pred. No. 1.2e-77;
Matches 537; Conservative   0; Mismatches 14;   Indels 0;   Gaps 0;

Qy   3016 tttgccaggggcatagatgttcacaaatgcagagttcgttctgactgagtgctcaactcaa 3075
           |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   551 TTTGCCAGAGGCATAGAGTGTTCACAATGCAGAGTTCGTTCTCGACTGGAGTGCTCACTCAA 492

Qy   3076 acgctggactatgaatcatataagtttaactgatgcgtctaggctgccgtgcatgtcga 3135
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   491 CCGCTGGACTATATATCATATAAGTTTAACTCAATGGCGTCTAGGCTCCGGTGCATGTCGA 432

Qy   3136 ctcctgcggtgcgggggctggctgtctggctggcggaggagctgctgcgcttccttcacag 3195
           ||||||||||||||| |||||||||||||||||||||||||||| |||||| |||||
Db   431 CTCCTCGGTGCGGGGGCTGGCTGGCTGGCTGGCGGCAGGAAGAGCTGCTGCGCCTTCCTTCACATG 372

Qy   3196 ctcttgtttcccagctgcttccctgggggatcagactgtgaaggacaggaagacagagatataa 3255
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371 CTCTTGTTTCCAACTGCTTCCCTGGGGGATCAGACTGTGAAGCAGAAGCAGGAGATATAA 312

Qy   3255 taaatatactgcatctcttttaagatgtgcaatttctgaggaaaacaataaattatgtt 3315
           |||||||||||||||| |||| ||||||||||||| |||||||||| |||||||||||
Db   311 TAAATATACCTGCATCTTTTTAAGATGTGCAATTTTATTCTGAGGAAAACATAAATTAATGTT 252

Qy   3316 ttgtatatatgactttaagagcccacattagttttatgattcaattgccaggttttta 3375
           |||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   251 TTGTATATATGCCTTTAAGAGCCCACATTAGGTTTTATGATTCATTGCCAGGTTTTTA 192

Qy   3376 aatgtttcacaaaactgtacggagacttcaactagaaataaaatggtgtaaataaagac 3435
           |||||||| ||||| |||||| ||||||||||||||||||||||||||||||||||
Db   191 AATGTTTTCCCAAACCTGTTACGGCGCCTTCAACTAGAAATAAAATGGTGTAAATAAAGGCC 132

Qy   3436 cttgctatctctaaattatggatgtgtaaagagattttgaaatgtttactttgattatttt 3495
           |||||||| | |||||||||||||||||||||| ||||||||||||||||||||||||
Db   131 CTTGCTATCTCTAAATTATGGGAGTGTAAAAGATTTGAAAATGTTTTGTACTTTGATTATTTT 72

(bottom right)

RESULT 15
AI393049/c
AI393049

```
Qy   3496 tatttcttatactctgtttctcttattgatatcttgcccacatttaaataaatgta 3555
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db     71 TATTTCTTATACTCTGTTTTTCTTTATTAATTGATATCTTGCCCATTTTAAATAAATGTA 12

Qy   3556 ctttttgaactt 3566
          ||||||||||||
Db     11 CTTTTTGAACTT 1
```

Search completed: June 4, 2002, 13:48:17
Job time: 9535 sec

OM nucleic - nucleic search, using sw model

   CenCore version 4.5
   Copyright (c) 1993 - 2000 Compugen Ltd.

Run on: June 4, 2002, 12:40:22 ; Search time 120.83 Seconds
   (without alignments)
   7277.729 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacaagtcgatacgcct.........gaacttagaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
   Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
   Maximum Match 100%
   Listing first 45 summaries

Database : Issued_patents_NA:*
   1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
   2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
   3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
   4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
   5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
   6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	%	DB ID	Description
1	110.6	3.1	397	3	US-09-253-691-3	Sequence 3, Appli
2	103	2.9	234	1	US-08-469-802B-3	Sequence 3, Appli
3	103	2.9	234	1	US-08-469-802B-3	Sequence 3, Appli
4	100.2	2.8	477	4	US-08-267-803B-3	Sequence 3, Appli
5	94.8	2.6	203	4	US-09-135-994-1	Sequence 1, Appli
6	92.6	2.6	543	6	US-09-043-303-7	Sequence 7, Appli
7	92.4	2.6	195	1	Patent No. 5273901	Patent No. 5273901
8	88.2	2.5	165	4	US-08-469-802B-2	Sequence 2, Appli
9	88.2	2.5	195	4	US-08-267-803B-2	Sequence 2, Appli
10	80.6	2.3	165	4	US-09-043-303-17	Sequence 17, Appl
11	80.6	2.3	10348	3	US-08-457-273B-41	Sequence 41, Appl
12	80.6	2.3	10348	3	US-08-556-419-13	Sequence 13, Appl
13	80.6	2.3	10348	4	US-09-041-886-14	Sequence 14, Appl
14	80.6	2.3	10366	1	US-08-246-982A-5	Sequence 5, Appl
15	80.6	2.3	10366	1	US-08-453-265-5	Sequence 5, Appli
16	78	2.2	154	4	US-08-469-802B-6	Sequence 6, Appli
17	78	2.2	154	2	US-08-267-803B-6	Sequence 6, Appli
18	77.2	2.2	533	6	5482709-5	Patent No. 5482709
19	77	2.2	3489	2	US-08-728-323A-1	Sequence 1, Appli
20	77	2.2	32207	2	US-08-770-379-20	Sequence 20, Appl
21	77	2.2	32207	4	US-08-757-669A-20	Sequence 20, Appl
22	76.6	2.1	32207	4	US-09-230-371A-20	Sequence 20, Appl
23	74.6	2.1	2190	4	US-09-625-188-19	Sequence 19, Appl
24	74.6	2.1	168	2	US-08-469-802B-4	Sequence 4, Appl
25	74.6	2.1	168	2	US-08-267-803B-4	Sequence 4, Appli
26	74.6	2.1	171	1	US-08-469-802B-5	Sequence 5, Appli
27	70.4	2.0	2214	3	US-08-864-038A-1	Sequence 1, Appli
c 28	70.4	2.0	3331	3	US-08-864-038A-2	Sequence 2, Appli
c 29	70.4	2.0	3331	3	US-08-864-038A-4	Sequence 4, Appli
c 30	69.4	1.9	688	4	US-08-998-416-915	Sequence 915, App
c 31	68.2	1.9	336	2	US-07-814-220-3	Sequence 3, Appli
32	68.2	1.9	336	2	US-07-812-421-3	Sequence 3, Appli
33	67.2	1.9	3292	1	US-07-814-964-12	Sequence 12, Appl
34	67.2	1.9	3292	1	US-08-258-442-12	Sequence 12, Appl
35	67.2	1.9	3292	1	US-08-328-809-7	Sequence 7, Appli
36	67.2	1.9	3292	5	PCT-US92-11107-12	Sequence 12, Appl
37	66.6	1.9	336	2	US-07-814-220-4	Sequence 4, Appl
c 38	66.6	1.9	336	2	US-07-814-220-4	Sequence 4, Appli
c 39	66.6	1.9	336	2	US-07-812-421-4	Sequence 4, Appli
c 40	66.6	1.9	336	2	US-07-812-421-4	Sequence 4, Appli
c 41	65.4	1.8	3366	2	US-08-469-803B-1	Sequence 1, Appli
42	65.4	1.8	3366	2	US-08-267-803B-1	Sequence 1, Appli
c 43	65	1.8	336	2	US-07-814-220-3	Sequence 3, Appli
c 44	65	1.8	336	2	US-07-812-421-3	Sequence 3, Appli
45	64	1.8	3376	1	US-08-320-559-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-253-691-3
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 3.1%; Score 110.6; DB 3; Length 397;
Best Local Similarity 63.2%; Pred. No. 2.5e-18;
Matches 170; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

```
Qy   1435 tttagccaagtgaaagtgactccagagacacacatgctcagacagcagcaggccag 1494
          ||||||    |||||| ||||| |||||||||||||||||||||||||||||  |||
Db    113 ttcagacagcagcaaaagcacagcagcagcagcagcagcagcagcagcagcag      172

Qy   1495 cagcagcagcagcagcagccggtgtttacaacctttcagcccagcgagataatgcagctccag 1554
          |||||||||||||||||| ||         ||||  ||||| |||||||| |||
Db    173 cagcagcagcagcagcagc|||||||||||          ||      ||| 232

Qy   1555 cagcagcagcagcagatctctcagcaaccttacccccagcagcagcgccgcatccattt 1614
          |||||||||||||| ||  | | ||| ||  ||| |||||||||||||||||||||||
Db    233 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagca 292

Qy   1615 tcacagcagcagcagcagcaaacagcagcaccacatcgcctcagcagcaatcagcttttt 1674
          ||||||||||||||||||||||||||||| || ||||||||||| |||||| ||||
Db    293 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 352

Qy   1675 ggacatgatccagcagtggagattccaga 1703
          |  |   ||| ||| |||||||| || |||
Db    353 cagcggggacctatcagcacagagttcaca 381
```

RESULT 2
US-08-469-802B-3
; Sequence 3, Application US/08469802B

; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

Query Match   2.98;   Score 103; DB 1; Length 234;
Best Local Similarity 66.4%;   Pred. No. 1.5e-16;
Matches 148; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy   1465 cacatgctacagcagcagcagcagcggcccagcagcagcagcagcaaccggtttacac 1524
          ||  ||||||||||||||||||||||||||||||||||||||||||||  ||
Db      4 CAGCAGCAGCAGCAGCAGCAGCCGCAGCAGCAGCCF/GCAGCAGCAGCAG 63

Qy   1525 cttcagccccagagataatgcagctccagcagcagcagcagcagcagatctctcagcaa 1584
          ||  ||||| ||||||||||||||||||||||||||||||||||| |||||
Db     64 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123

Qy   1585 ccttacccccagcagccgccgccatccaatttcacagcagcagcaacagcagcagcca 1644
          |||||||||||||||||  |||  ||||
Db    124 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183

Qy   1645 ccaccatcgcctcagcagcatcaagcttttggacatgatccag 1687
          |  |||||||||| ||||  |||
Db    184 CAGCAGCAGCAGCAGCAGCACCTCAGCAGGGCTCCG 226
```

RESULT 3
US-08-267-803B-3
; Sequence 3, Application US/0826780803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.

; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55458-1415
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/267,803B
; FILING DATE: 28-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McCormack, Myra H.
; REGISTRATION NUMBER: 36,602
; REFERENCE/DOCKET NUMBER: 110.00030120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1228
; TELEFAX: 612-305-1217
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-267-803B-3

Query Match   2.98;   Score 103; DB 2; Length 234;
Best Local Similarity 66.4%;   Pred. No. 1.5e-16;
Matches 148; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy   1465 cacatgctacagcagcagcagcagcggcccagcagcagcagcagcaaccggtttacac 1524
          ||  ||||||||||||||||||||||||||||||||||||||||||||  ||
Db      4 CAGCAGCAGCAGCAGCAGCAGCCGCAGCAGCAGCAGCAGCGACGCAGCAGCACCAG 63

Qy   1525 cttcagccccagagataatgcagctccagcagcagcagcagcagcagatctctcagcaa 1584
          ||  ||||| |||||||||||||||||||||||||||||||||| |||||
Db     64 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 123

Qy   1585 ccttacccccagcagccgccgccatccaatttcacagcagcagcaacagcagcagcca 1644
          |||||||||||||||||  |||  ||||
Db    124 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 183

Qy   1645 ccaccatcgcctcagcagcatcaagcttttggacatgatccag 1687
          |  |||||||||| ||||  |||
Db    184 CAGCAGCAGCAGCAGCAGCACCTCAGCAGGGCTCCG 226
```

RESULT 4
US-09-135-994-1
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Ranum et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1

```
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-1

Query Match     2.8%;  Score 100.2;  DB 4;  Length 477;
Best Local Similarity   69.9%;  Pred. No. 1.1e-15;
Matches  135;  Conservative   0;  Mismatches  58;  Indels   0;  Gaps   0;

Qy   1477 cagcagcagcagcagccagcgagcgcagcagcaaccggtttacacttcagcccag 1536
          | |||||||| ||||||||||||||||||| | ||            ||| |||
Db    124 cggcagcagcagcagcagcagcagcagcagcagatctctcagcaacttaccccag 183

Qy   1537 cagataatgcagctccagcagcagcagcagcagcagatctctcagcaacttaccccag 1596
          |||          ||||||||||||||||||||||||||||||||||  ||  |||
Db    124 cggcagcagcagcagcagcagcagcagcagcagatctctcagcaacttaccccag 183

Qy   1597 cagccgccgatccatttcacacagcagcagcagcaacagcagcaacatcgcct 1656
          ||||   |||  ||| |||  |||||||||||||  ||||||||||    ||
Db    244 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcc 303

Qy   1657 cagcagcagcatcagc 1669
          |||| |||||| ||||
Db    304 cagcagcagcagcagc 316
```

RESULT   5
US-09-043-303-7
; Sequence 7, Application US/09043303
; Patent. No. 6251589
; GENERAL INFORMATION:
; APPLICANT: TSUI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 203
; TYPE: DNA
; ORGANISM: p-2093 plasmid
US-09-043-303-7

Query Match   2.6%; Score 94.8; DB 4; Length 203;
Best Local Similarity   70.8%; Pred. No. 1.5e-14;
Matches 126; Conservative   0; Mismatches 52; Indels   0; Gaps   0;

```
Qy   1465 cacatgctacagcagcagcagcagcccagcagcagcagcaccggtttacac 1524
          |||     ||||||||||||| |||||||||||||||| ||
Db      4 caccagccaacagcaacagcagcagcagcagcagcagcagcagcagcagcag 63

Qy   1525 cttcagcccagcagcaataatgcagctccagcagcagcagatctctcagcaa 1584
          ||| ||||||||||||                   |||| ||||
Db     64 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 123

Qy   1585 ccttaccccagcagcacgcgccacatccatttcacacagcagcaacagcagc 1642
          || |||||||||||||||            ||||||
Db    124 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 181
```

RESULT   6
5273901-6
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.;STRAUSBERG, ROBERT L.;WILSON,
; SUSAN D.;POPE, SHARON H.;STRAUSBERG, SUSAN L.;RUFF,MICHAEL D.;
; AUGUSTINE, PATRICIA C.;DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
;SPOROZOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
;SEQ ID NO:6:
; LENGTH:543
5273901-6

Query Match   2.8%; Score 92.6; DB 6; Length 543;
Best Local Similarity   56.9%; Pred. No. 9.1e-14;
Matches 170; Conservative   0; Mismatches 129; Indels   0; Gaps   0;

```
Qy   1369 acgaaaaatttagaacacagcaggtgaattcacacagccagcggggaccatacaaatgccaatgca 1428
          || |     || |||       ||     |||   |||||||          |||      ||
Db      5 aagaagacaaaagagagagagacaaaagagaagagagagagaagaagaagaagagaagaga 64

Qy   1429 gtgctgtttagccaagtgaaagtgactccagacacacatgctacagcagcagcagcag 1488
          ||      ||       || ||          |||                ||
Db     65 gaggagagaagagaagaagaagagcgggcagcacctgcagcagcaacagcag 124

Qy   1489 gccagcagcagcagcagcagcaccggtttacacttcagcccaccccagcagccagc 1548
          |||||||||||||||||||            |||    |||| |||
Db    125 cagcagcctgcagcagcaacagcagcagcagcagcagcagcagcagcagctacagcaa 184

Qy   1549 ctccagcagcagcagcagatctctcagcaacttaccccagcagcagcgcgcat 1608
          ||||||||||||||     |||| ||||          |||||||  |||
Db    185 caccagctgc-gctgcagcaaccagcagctgcagcagcagcagcagcagccgcat 244

Qy   1609 ccattttcacagcagcagcggcaacagcagccaccatcgcctcagcagcatca 1667
          |||||||||||||||||        ||||||||  |||| |||||| ||
Db    245 caggagctgcagcagcaccagcagctgcagcagcagccgcagcagcagcaggagcagcagcaa 303
```

RESULT   7
US-08-469-802B-2
; Sequence 2, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Ranum, Laura P.W.
; APPLICANT: Chung, Ming-yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

NAME: Mueting, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00030101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1225
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-469-802B-2

Query Match   2.68;   Score 92.4; DB 1; Length 195;
Best Local Similarity   67.9%; Pred. No. 6e-14;
Matches 129; Conservative   0; Mismatches   61; Indels   0; Gaps   0;

```
Qy   1474 cagcagcagcagcgcccagcagcagcagcgaacccggtttttacaccttcagcc 1533
          ||||| ||||||||      |||||||||||          |||| |||  |||
Db      1 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAG   60

Qy   1534 cagcagataatgcagctccagcagcagcagagatctctcagcaaacttaccc 1593
          |      |||   |||   |||||||||||||        |||| |||  |||
Db     61 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAG   120

Qy   1594 cagcagccgccgcatccatttcacagcagcagcagcagcaacaccatcg 1653
          |||||| | |  |||          ||||||||||||| ||||
Db    121 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGG   180

Qy   1654 cctcagcagc 1663
          |||  | ||
Db    181 GCTCCGGCGC  190
```

RESULT 8
US-08-267-803B-2
Sequence 2, Application US/0826780 3B
Patent No. 5834183
GENERAL INFORMATION:
APPLICANT: Orr, Harry T.
APPLICANT: Ranum, Laura P.W.
APPLICANT: Chung, Ming-Yi
APPLICANT: Zoghbi, Huda Y.
TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
Patent No. 5834183
TITLE OF INVENTION: Type 1 and Method for Diagnosis
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
STREET: P.O. Box 581415
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55458-1415
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267,803B
FILING DATE: 28-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McCormack, Myra H.
REGISTRATION NUMBER: 36,602
REFERENCE/DOCKET NUMBER: 110.00030120
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-267-803B-2

Query Match   2.68;   Score 92.4; DB 2; Length 195;
Best Local Similarity   67.9%; Pred. No. 6e-14;
Matches 129; Conservative   0; Mismatches   61; Indels   0; Gaps   0;

```
Qy   1474 cagcagcagcagcgcccagcagcagcagcgaacccggtttttacaccttcagcc 1533
          ||||| ||||||||      |||||||||||          |||| |||  |||
Db      1 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAG   60

Qy   1534 cagcagataatgcagctccagcagcagcagagatctctcagcaaacttaccc 1593
          |      |||   |||   |||||||||||||        |||| |||  |||
Db     61 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAG   120

Qy   1594 cagcagccgccgcatccatttcacagcagcagcagcaacaccatcg 1653
          |||||| | |  |||          |||||||||||| ||||
Db    121 CAGCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCTCAGCAGG   180

Qy   1654 cctcagcagc 1663
          |||  | ||
Db    181 GCTCCGGCGC  190
```

RESULT 9
US-09-043-303-17
Sequence 17, Application US/09043303
Patent No. 6251589
GENERAL INFORMATION:
APPLICANT: TSUJI, Shoji
APPLICANT: SANPEI, Kazujiro
TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Ataxia Type 2 and
TITLE OF INVENTION: Primers Therefor
FILE REFERENCE: 0760-0241P
CURRENT APPLICATION NUMBER: US/09/043,303
CURRENT FILING DATE: 1998-05-18
EARLIER APPLICATION NUMBER: PCT/JP96/01999
EARLIER FILING DATE: 1996-07-18
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 165
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Probe
US-09-043-303-17

Query Match   2.58;   Score 88.2; DB 4; Length 165;
Best Local Similarity   70.9%; Pred. No. 6.1e-13;
Matches 117; Conservative   0; Mismatches   48; Indels   0; Gaps   0;

```
Qy   1474 cagcagcagcagcgcccagcagcagcagcagcagcaccaccggtttttacaccttcagcc 1533
          ||||| |||||||      ||||||||||||||||||          |||| |||  |||
Db      1 cagcagcagcagcagcagcagcagcagcagcagcagatctctcagcagcagcagcagcag   60

Qy   1534 cagcagataatgcagctccagcagcagcagagatctctcagcaaacttaccc 1593
          |      |||   |||   ||||||||||||||        |||| |||  |||
Db     61 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag   120

Qy   1594 cagcagccgccgcatccatttcacagcagcagcagcagcagcagcaacag 1638
          |||||| | |  |||          |||||||||||||||||| |||
Db    121 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 165
```

RESULT 10
US-08-457-273B-41
; Sequence 41, Application US/08457273B
; Patent No. 5849995
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael
; APPLICANT: Lin, Biaoyang
; APPLICANT: Nasir, Jamal
; TITLE OF INVENTION: Mouse Model for Huntington's Disease and
; TITLE OF INVENTION: Related DNA Sequences
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Virginia Bennett
; STREET: PO Box 37428
; CITY: Raleigh
; STATE: No. 5849995th Carolina
; COUNTRY: US
; ZIP: 27627
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,273B
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Virginia C.
; REGISTRATION NUMBER: 37,092
; REFERENCE/DOCKET NUMBER: 3477-85A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-854-1400
; TELEFAX: 919-854-1401
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10348 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-457-273B-41

Query Match   2.3%; Score 80.6; DB 2; Length 10348;
Best Local Similarity   62.8%; Pred. No. 4.1e-10;
Matches 125; Conservative   0; Mismatches 74; Indels   0; Gaps   0;

```
Qy 1459 gagacacacatgctacagcagcagcagcagcccagcagcagcagcagcagcagcagcacccggtt 1518
         |||||||||||||||||||||||||||||||||||||||||||||
Db  349 gagtccctcaagtccttccagcagcgcagcgagcagcagcagcagcagcagcagcagcag 408

Qy 1519 ttacacttcagcccagcagatataatgcagctccagcagcagcagcagcagcagatctct 1578
         ||||||  ||||| ||||| |||||||||||||||||||||||||||||||
Db  409 cagcagcagcagcaacaccgccaccccccccccccccccccgcccctcct 468

Qy 1579 cagcaaccttaccccagcagcagcgcgcgccatccatttcacagcagcagcagcaacag 1638
         ||||| |||  ||| || |   |||||| || |||||||||| || ||| |
Db  469 cagcttccctcagcccgccgccgccgccgaggcacagcagcgctgctgctcagccgccagcccccccg 528

Qy 1639 cagcaccaccaatcgcctc 1657
         | ||| ||||| ||||||
Db  529 ccgccgccccgccgccac 547
```

RESULT 11
US-08-556-419-13
; Sequence 13, Application US/08556419C
; Patent No. 6093549
; GENERAL INFORMATION:
; APPLICANT: Ross, Christopher
; APPLICANT: Li, Xiao-Jiang
; APPLICANT: Li, Shi-Hua
; APPLICANT: Sharp, Alan
; APPLICANT: Lanahan, Anthony
; APPLICANT: Worley, Paul
; APPLICANT: Snyder, Solomon
; TITLE OF INVENTION: Huntingtin-associated protein
; FILE REFERENCE: 01107.52271
; CURRENT APPLICATION NUMBER: US/08/556,419C
; CURRENT FILING DATE: 1995-11-09
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 10348
; TYPE: DNA
; ORGANISM: Homo sapiens
US-08-556-419-13

Query Match   2.38; Score 80.6; DB 3; Length 10348;
Best Local Similarity   62.8%; Pred. No. 4.1e-10;
Matches 125; Conservative   0; Mismatches 74; Indels   0; Gaps   0;

```
Qy 1459 gagacacacatgctacagcagcagcagcagcccagcagcagcagcagcagcagcagcacccggtt 1518
         |||| | ||                  ||||||||||||||||||||||||||||||||
Db  349 gagtccctcaagtccttccagcagcgcagcgagcagcagcagcagcagcagcagcagcag 408

Qy 1519 ttacacttcagcccagcagatataatgcagctccagcagcagcagcagcagcagatctct 1578
         ||| || |||||                |||||||||||||||||||||||||||
Db  409 cagcagcagcagcaacaccgccaccccccccccccccccccgcccctcct 468

Qy 1579 cagcaaccttaccccagcagcagcgcgcgccatccatttcacagcagcagcagcaacag 1638
         ||||| ||  || || |   ||||| || |||||||||| || ||| |
Db  469 cagcttccctcagcccgccgccgccgccgaggcacagcagcgctgctgctcagccgccagcccccccg 528

Qy 1639 cagcaccaccaatcgcctc 1657
         | ||| ||||| ||||||
Db  529 ccgccgccccgccgccac 547
```

RESULT 12
US-09-041-886-14
; Sequence 14, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Shahrooz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:

```
;  LENGTH: 10348 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: DNA (genomic)
;  FEATURE:
;    NAME/KEY: CDS
;    LOCATION: 316..9748
US-09-041-886-14
```

```
Query Match      2.3%;  Score 80.6;  DB 4;  Length 10348;
Best Local Similarity 62.8%;  Pred. No. 4.1e-10;
Matches  125;  Conservative   0;  Mismatches   74;  Indels   0;  Gaps   0;
US-08-246-982A-5

Qy  1459 gagacacacatgctacagcagcagcagcagcgccagcagcagcagcagcaccggtt 1518
         |||||||||||||||||||||||          |||||||||||||||||||||||
Db   349 GAGTCCCTCAAGTCCTTCAGCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy  1519 ttacaccttcagcccagcagataatgcagctccagcagcagcagcagcagatctct 1578
         ||||||||||||||| ||| |||  |||||||||||||||||| |||| ||
Db   409 CAGCAGCAGCAGCAGCAACAGCCCCGCCGCCACCGCCGCCGCCGCCGCCGCCTCCT 468

Qy  1579 cagcaaaccttacccccagcagcgccgccgccgccatccatttcacagcagcaacag 1638
         |||  || |||||||||||||||||||||||||| ||| |||  ||||| ||
Db   469 CAGCTTCCTCAGCCGCCGCCGGCCAGCACAGCCGCTCCTCCCTCAGCCGCCGCCCCG 528

Qy  1639 caggccaccaccatcgcctc 1657
         ||| ||| || |||| ||||
Db   529 CCGCCGCCGCCGCCGCCCAC 547
```

```
RESULT 13
US-08-246-982A-5
; Sequence 5, Application US/0824698 2A
; Patent No. 5686288
;  GENERAL INFORMATION:
;  APPLICANT: MacDonald, Marcy E.
;  APPLICANT: Ambrose, Christine M.
;  APPLICANT: Duyao, Mabel P.
;  APPLICANT: Gusella, James F.
;  TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
;  NUMBER OF SEQUENCES: 25
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;    STREET: 1100 New York Avenue
;    CITY: Washington
;    STATE: D.C.
;    COUNTRY: U.S.A.
;    ZIP: 20005
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: Patentin Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/246,982A
;    FILING DATE: May 20, 1994
;    CLASSIFICATION: 435
;  ATTORNEY/AGENT INFORMATION:
;    NAME: Goldstein, Jorge, A.
;    REGISTRATION NUMBER: 29,021
;    REFERENCE/DOCKET NUMBER: 0609.3880002
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (202) 371-2600
;    TELEFAX: (202) 371-2540
;  INFORMATION FOR SEQ ID NO: 5:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 10366 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;  FEATURE:
```

```
;  NAME/KEY: CDS
;  LOCATION: 316..9748
US-08-246-982A-5
```

```
Query Match      2.3%;  Score 80.6;  DB 1;  Length 10366;
Best Local Similarity 62.8%;  Pred. No. 4.1e-10;
Matches  125;  Conservative   0;  Mismatches   74;  Indels   0;  Gaps   0;

Qy  1459 gagacacacatgctacagcagcagcagcagcgccagcagcagcagcagcaccggtt 1518
         |||||||||||||||||||||||          |||||||||||||||||||||||
Db   349 GAGTCCCTCAAGTCCTTCAGCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy  1519 ttacaccttcagcccagcagataatgcagctccagcagcagcagcagcagatctct 1578
         ||||||||||||||| ||| |||  |||||||||||||||||| |||| ||
Db   409 CAGCAGCAGCAGCAGCAACAGCCCCGCCGCCACCGCCGCCGCCGCCGCCGCCTCCT 468

Qy  1579 cagcaaaccttacccccagcagcgccgccgccgccatccatttcacagcagcaacag 1638
         |||  || |||||||||||||||||||||||||| ||| |||  ||||| ||
Db   469 CAGCTTCCTCAGCCGCCGCCGGCCAGCACAGCCGCTCCTCCCTCAGCCGCCGCCCCG 528

Qy  1639 caggccaccaccatcgcctc 1657
         ||| ||| || |||| ||||
Db   529 CCGCCGCCGCCGCCGCCCAC 547
```

```
RESULT 14
US-08-453-265-5
; Sequence 5, Application US/0845265
; Patent No. 5693757
;  GENERAL INFORMATION:
;  APPLICANT: MacDonald, Marcy E.
;  APPLICANT: Ambrose, Christine M.
;  APPLICANT: Duyao, Mabel P.
;  APPLICANT: Gusella, James F.
;  TITLE OF INVENTION: Huntingtin DNA, Protein And Uses Thereof
;  NUMBER OF SEQUENCES: 25
;  CORRESPONDENCE ADDRESS:
;    ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;    STREET: 1100 New York Avenue
;    CITY: Washington
;    STATE: D.C.
;    COUNTRY: U.S.A.
;    ZIP: 20005
;  COMPUTER READABLE FORM:
;    MEDIUM TYPE: Floppy disk
;    COMPUTER: IBM PC compatible
;    OPERATING SYSTEM: PC-DOS/MS-DOS
;    SOFTWARE: Patentin Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;    APPLICATION NUMBER: US/08/453,265
;    FILING DATE: 30-MAY-1995
;    CLASSIFICATION: 514
;  ATTORNEY/AGENT INFORMATION:
;    NAME: Ludwig, Steven R.
;    REGISTRATION NUMBER: 36,203
;    REFERENCE/DOCKET NUMBER: 0609.3880003
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (202) 371-2600
;    TELEFAX: (202) 371-2540
;  INFORMATION FOR SEQ ID NO: 5:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH: 10366 base pairs
;    TYPE: nucleic acid
;    STRANDEDNESS: single
;    TOPOLOGY: linear
;  FEATURE:
```

```
Query Match      2.3%;  Score 80.6;  DB 1;  Length 10366;
```

```
Best Local Similarity   62.8%;  Pred. No. 4.1e-10;
Matches 125;  Conservative   0;  Mismatches   74;  Indels   0;  Gaps   0;

Qy   1459 gagacacacatgctacagagcaggcaagcaggcaagcaccggtt 1518
          ||||-|-|-|-||                       ||||||||||||||
Db    349 CACTCCCTCACGTCCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 408

Qy   1519 ttacacttcagcccagcagataatgcagcagcagcagcagagatct 1578
          |-|||  ||-|||||||||||||||||||||||||||||||||||
Db    409 CAGCAGCAGCAGCAGCAGCAGCAACACGCGGCCACCGCCGCCGCCTCCT 468

Qy   1579 cagcaaacttaccccagcagcccgcccggatccatttcacagcagcagcagcagcaacag 1638
          ||||    ||  |  |-||-|-|||-|-|||-||-|-|-|||-|-||-|-|-||
Db    469 CAGCTTCCTCACCCGCCGCCAGCCACCCGCTTCCTTCAGCCGCAGCAGCCCCCCC 528

Qy   1639 cagccaacaacatcagcctc 1657
          | |-||-|| |||-|
Db    529 CCGCCGCCCCCCCCCCCCAC 547
```

```
RESULT 15
US-08-469-802B-6
; Sequence 6, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
;   APPLICANT:  Orr, Harry T.
;   APPLICANT:  Ranum, Laura P.W.
;   APPLICANT:  Chung, Ming-yi
;   APPLICANT:  Zoghbi, Huda Y.
;   TITLE OF INVENTION:  Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
;   TITLE OF INVENTION:  Type 1 and Method for Diagnosis
;   NUMBER OF SEQUENCES:  47
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Mueting, Raasch, Gebhardt & Schwappach, P.A.
;     STREET:  119 No. 5741645th Fourth Street, Suite 203
;     CITY:  Minneapolis
;     STATE:  MN
;     COUNTRY:  USA
;     ZIP:  55401
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/469,802B
;     FILING DATE:  06-JUN-1995
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Mueting, Ann M.
;     REGISTRATION NUMBER:  33,977
;     REFERENCE/DOCKET NUMBER:  110.00030101
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  612-305-1217
;     TELEFAX:  612-305-1225
; INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  154 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  DNA
US-08-469-802B-6

Query Match            2.2%;  Score 78;  DB 1;  Length 154;
Best Local Similarity  70.0%;  Pred. No. 2.1e-10;
Matches 105;  Conservative   0;  Mismatches  45;  Indels   0;  Gaps   0;

Qy   1493 agcagcaagcaagcaaccgttcacacctcagcccagagataatgcagctcc 1552
          |||||||||||||||||||||||| |  |||      ||||   ||||
Db      3 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGC 62
```

```
Qy   1553 agcaagcagcagcagcagcagatctctcagcaaccttaccccccagcagcaggccggccagccgccgccatccat 1612
          ||||||||||||||||||||||     |||| ||  |  |||||||||||| |-|  |||||||||||||||||||||| ||||  |
Db     63 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 122

Qy   1613 tttcacagcagcagcagcagcaacagcagc 1642
          ||||||||||||||||||||||| |||||||
Db    123 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCCACC 152
```

Search completed: June 4, 2002, 15:36:56
Job time: 10594 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run On: June 4, 2002, 12:45:52 ; Search time 529 Seconds
       (without alignments)
       11619.194 Million cell updates/sec

Title: US-09-664-641-10
Perfect score: 3580
Sequence: 1 catagacagtcgatacgcct...........gaacttagaaaaaaaaaa 3580

Scoring table: IDENTITY_NUC
      Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Post-processing: Minimum Match 0%
      Maximum Match 100%
      listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Database :   N_Geneseq_032802:*
   1: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
   2: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
   3: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
   4: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
   5: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
   6: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
   7: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
   8: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
   9: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
  10: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
  11: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
  12: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
  13: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
  14: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
  15: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
  16: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
  17: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
  18: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
  19: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
  20: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
  21: /SIDS5/gcgdata/geneseq/geneseqn-cmbl/NA2000.DAT:*
  22: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
  23: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
  24: /SIDS5/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	3519	98.3	3519	22	AAS44722	Human full-length
2	331.2	9.3	335	21	AAC27011	Human secreted pro
3	204.6	5.7	690	22	AAH70661	Human cervical can
4	186.8	5.2	504	22	AAH71865	Human cervical can
5	186.8	5.2	579	22	AAH68998	Human cervical can
6	162	4.5	467	22	AAI85135	Human polynucleoti
7	135.6	3.8	150	19	AAX11992	Human biallelic po
c	8	125	3.5	344	22	AAS44894	Human contig polyn
9	110.6	3.1	397	20	AAX89891	Spinocerebellar at

---

c	10	105.6	2.9	5120	22	AAC84677	DNA sequence of hu
c	11	105.2	2.9	1037	21	AAS59242	Exons E, C and A o
c	12	105.2	2.9	1159	21	AAS59240	An EcoRI fragment
c	13	105.2	2.9	1472	21	AAS59241	Exons D, C, B and
	14	103	2.9	234	16	AAQ84832	Spinocerebellar at
	15	100.2	2.8	477	21	AAX44307	Human SCA7 genomic
	16	94.8	2.8	486	22	AAV17226	Glutamine rich reg
	17	94.8	2.6	203	19	AAV30271	SCA2 gene CAG repe
	18	94.2	2.6	6794	22	AAF75507	Polyglutamine trac
	19	93.2	2.6	543	19	AAZ45597	cDNA sequence of h
	20	92.6	2.6	543	13	AAQ23092	Antigen tc-7a gene
	21	92.4	2.6	195	16	AAQ84831	Spinocerebellar at
	22	89.6	2.5	567	21	AAA29550	HIV codon altered
c	23	89.4	2.5	1400	23	AAS83195	DNA encoding novel
c	24	88.2	2.5	165	19	AAV30274	Glutamine rich reg
c	25	86.6	2.4	6273	23	ABL15731	Drosophila melanog
	26	86.6	2.4	10855	23	ABL15730	Drosophila melanog
	27	86	2.4	4335	23	AAS79854	Drosophila melanog
	28	85	2.4	9732	23	AAS79690	DNA encoding novel
c	29	84.8	2.4	3263	23	ABL10403	Drosophila melanog
	30	84.8	2.4	16962	23	ABL10402	Drosophila melanog
	31	84.4	2.4	4409	23	ABL17895	Drosophila melanog
	32	83.4	2.3	5021	23	ABL11783	Drosophila melanog
	33	83.4	2.3	7994	23	ABL17894	Drosophila melanog
	34	83.4	2.3	7614	23	ABL11782	Drosophila melanog
	35	83.2	2.3	10910	23	ABL03131	Drosophila melanog
	36	83	2.3	2215	24	ABI99688	Drosophila melanog
	37	83	2.3	6558	23	AAZ45602	Mouse ischemic co
	38	82.8	2.3	2672	23	ABL08857	cDNA sequence of a
	39	82.8	2.3	4810	23	ABL08856	Drosophila melanog
c	40	82.4	2.3	14704	13	AAQ20685	Drosophila melanog
	41	82.2	2.3	10348	22	ABA08714	PKS 741 insert con
	42	82	2.3	629	21	AAF08420	Human Huntington's
	43	81.4	2.3	1713	23	ABL18459	Fusarium venenatum
	44	81.4	2.3	5157	23	ABL18458	Drosophila melanog
c	45	81.2	2.3	3791	23	ABL04336	Drosophila melanog

ALIGNMENTS

RESULT 1
AAS44722

   AAS44722 standard; DNA; 3519 BP.

XX
AC   AAS44722;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human full-length polynucleotide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; ds;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO20016483A-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.

PR   14-JUL-2000; 2000US-0216807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589962/66.
DR   P-PSDB; AAU278822.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection -
XX
PS   claim 1: SEQ ID No 147; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3519 BP; 996 A; 783 C; 839 G; 901 T; 0 other;

Query Match   98.38%; Score 3519; DB 22; Length 3519;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3519; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   62 agtgcaagagaaggggggggggtgctgaaggggccccccgggtctggtctgggggaactc 121
         |||||||||||||||| |||| |||||||||||||||||||||||||| ||||| |||||
Db   1 agtgcaagagaagggggcggggctgctgaaggggccccccgggtctggt.tgggcgggaactc 60

Qy   122 cgggacgggcgcccgcaggggctacggagtctctcgcccctcaaatcacctcta 181
         |||||||||||| ||||||||||||||| ||||||||||||||||||||||||||
Db   61 cgggacgggcgcgcaggggctacggatctctcggcccctcaaatcacctcta 120

Qy   182 ggcacaggttattcagcttctcaaggctggaaaaagaggaagttccctacaaatgcact 241
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ggcacaggttattcagcttctcaaggctggaaaaagaggaagttcctacaaatgcact 180

Qy   242 agcctcacacaataatccaagaagatgggagacaaggctccggaaat 301
         |||||||||||||||||||||||||||||||||||||||||||||||
Db   181 agcctcacacaataatccaagaaggtgggagaacaaggctcgggaagt 240

Qy   302 ctttgacttacctgtgtaaagccttgggtgtaagcttctgtcgttcagtgtggaactct 361
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 cttgacttacctgttgtaaagccttgggtgtaagctctgtcgttcagtgtggaactct 300

Qy   362 tctgtccagtaaatggttttctccagaatcatgtcagagaatttttttgaatcactcctg 421
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 tctgtccagtaaatggttttctccagaatcatgtcagagaattttttttgaatcactccctg 360

Qy   422 ccttctccaggatgttgttgtacaagctgggagctctttgtgagtctttccagagctccc 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ccttctccaggtgttgatacaagctgggagctctttgtgagtctttccagagctcccc 420

Qy   482 aaggagaaggtaggaaaggagctgtccagcagcaagaagtctggggaagaagacacatctgc 541
         ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   421 aaggagaaggtaggaaaggagctgtccagcagcaagaagtctggggaagaagacagac.tcatctgc 480

Qy   542 ctctttctgacccggttattgcaggctgaggcctcgtgttgttgaatgtctgggtgtcat 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ctctttctgacccggttattgcaggctgaggcctcgtgttgttgaatgtctgggttgtcat 540

Qy   602 ctgaagacaagaaagtgccctgtgggctttggttacgttctatgtggtgtcat 661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ctgaagacaagaaagtgccctgtgggctttggttacgttctatgtggtgtcat 600

Qy   662 ccctcaataagaaatgacgacgcatttgattgttccagagccaaaggggggggaaatacgaat 721
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 ccctcaataagaaatgacgcatttgattgttccagagccaaaggggggaaataacgaat 660

Qy   722 gtgcttaaggagcaagatattaaaattgtgactcctgactcggttctggattggctgtat 781
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 gtgcttaaggagcaagtattaaaattgtgactcctgactcggttctggattggcgtat 720

Qy   782 caggaagaaaacaaaagaacgacgaagcaattcatcatcctcgtctgatttatttatgaagagg 841
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 caggaagaaaacaaaagacgaagcaattcatcctcgtctgattattgaagagg 780

Qy   842 aagaagaggaagaggaagagtagaaaatgagaggaacaaaatttctcagaatg 901
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 aagaagaggaagaggaagagtagaaaatgagaggaacaaagatctcagaatg 840

Qy   902 agggtagtacagatgaagtcaagcctccagcctccagagaaggtctcctcagtg 961
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 agggtagtacagatgaagtcaagcctccagcctccagagaaggtctcctcagtg 900

Qy   962 aacaggcagtttcacctaaatccacaactgaaaaatctaaagcgaatttaatgttgatg 1021
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 aacagcagtttcacctaaatccacaactgaaaaatctaaagggaattaatgttgatg 960

Qy 1022 atctttcagattcatcacggaaaaacagagagaaaatttaaactgaactgtccagaag 1081
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 atctttcagattcatcacggaaaaacagagagaaaatttaaactgaactgtccagaag 1020

Qy 1082 tcccacagttagctcgcagcaaacgctgctcagggaaggagggctgggttgatta 1141
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 tcccacagttagctgcgacagcaaacgctgctcagggaaggagggctgggttgatta 1080

Qy 1142 acttgtgtgccaatgtcccacccgtccagtaacattttgccccctgagtccgggggta 1201
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 acttgtgtgccaatgtcccacccgtccagtaacatttgccccctgagtccgggta 1140

Qy 1202 atttaattggctgtcggacaaaacctccaaagttctgaaagatcagaaatgatagctacct 1261
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1141 atttaattggctgtcggacaaaacctccaaagttctgaaagatcagaaatgatagctact 1200

Qy 1262 gggttccagctgctacggacactgaggaaatattactaatgctacattcagcagatga 1321
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1201 gggttccagctgctacggacactgaggaaatattactaatgctacattcagcagatga 1260

Qy 1322 accggtccatcaaatgtagcacatatcctacagaactcttcagcacctacgaaaaatttag 1381
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1261 accggtccatcaaatgtagcacatatcctacagaactcttcagcacctacgaaaaatttag 1320

Qy 1382 aacagcaggtgaatcacagcagagacaatacaaatgccaatgcagtcagtttagc 1441
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1321 aacagcaggtgaatcacagcagagacaatacaaatgccaatgcagtcagttagc 1380

Qy 1442 aagtgaaagtgactccagagacacacatgctacagcagcagcccagcagcagagc 1501
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1381 aagtgaaagtgactccagagacacacatgctacagcagcaccccagcagcagagc 1440

Qy 1502 agcagcagcaccccggtttacacttcagcccagcagataagtgcagctccagcagcagc 1561

```
Db   1441  agcagcagcagcaccggtttacacttcagcccagcagcagtaatgcagctcagcagcagc
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1562  agcagcagcagatctctcagcagccttaccccagccagcagcagcagccagccattttcacagc 1621
Db   1501  agcagcagcagatctctcagcagccttaccccagccagcagcagcagccagccatcctcacagc 1560

Qy   1622  agcagcagcagccaccacatcagctcagcagcacatcagctttggacatg 1681
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  agcagcagcagccaccacatcagctcagcagcacatcagctttggacatg 1620

Db   1621  atccagcagtcggaatccagaaaaagcttcttattgggatgtgtgttgcaattgcgg 1680
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1682  atccagcagtcggaagttccagaaaaagcttcttattgggatgtgtgttgcaattgcgg 1741

Qy   1742  attatccagaggacgtgtcgataagcaacctgctggccacctggaaaagataatccagg 1801
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  attatccagaggacgtgtcgataagcaacctgctggccacctggaaaagataatccagg 1740

Db   1741  cacatgcggcacttgtgacccacccacttcacggcacctctcctgtgaga 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1802  cacatgcggcacttgtgacccacccacttcacggcacctctcctgtgaga 1861

Qy   1862  gtcaagtcagcacgcagcgcgtatgcacagcaataagcaagaaaagagagatgtgttactgcac 1921
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  gtcaagtcagcacgcagcgcgtatgcacagcaataagcaagaaaagagagatgtgttactgcac 1860

Db   1861  actggttaaacacagtcttaaaaaaaaaataagaagagaagaggataatccagg 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1922  actggttaaacacagtcttataaaaaaaatggtacgcgccgagcccttcact 1981

Qy   1982  tcccagtggcctccacccacccagaggaaagcaaccacatttattcttgtaactg 2041
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  tcccagtggccttcccacccagaggaaagcccatgttcacgcacatattattcttgtgactg 1980

Db   1981  gatttgttgataglgacagagaatgacctaaaattaaatgtcaagtgccaaat 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2042  gattttgttgataglgacagagaatgacctaaaattaattaggctatttcgctacacggaat 2101

Qy   2102  atacgggttatctatgccgagcacaacacagtctcatctgtaaaat 2161
           |||||||||||||||||||||||||||||||||||||||||||||||
Db   2041  atacgggtatctatgccgagcacaacacagtcctcatctgtaaaataccag 2100

Db   2101  agtatgaaaaagccaagaagtggaggaatacccctgtctcaacgcccagtgg 2160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2162  agtatgaaaaagccaagaagtggaggaatacccctgtctcaacgcccagtggcttggcgaca 2221

Qy   2222  ttcttctgggaaacttgagcgctcaggcactcaagtcagtagtcactcacacggatca 2281
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161  ttcttctgggaaacttgagcgctcaggcactcaagtcagtagtcactcacacggatca 2220

Db   2221  gtctgcaggatccattgccctcaccaccacagccattagtttaaatctttagaatgcttgga 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2282  gtctgcaggatccattgccctcaccaccacagccattagtttaaatctttagaatgcttgga 2341

Qy   2342  gaattccctaaaagtgtctgcagagttgatgaagtataagactacctcccaaactga 2401
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281  gagttccctaaaagtgtctgcagagttgatgaagtataagactacctcccaaactga 2340

Db   2401  caccccactaaaaagctaactccagaatgacccctttgtgcttcactggattcg 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2462  cacctccactaaaaagctaactccagaatgacccctttgtgcttcactggattcg 2521

Qy   2402  aacagagaatagctaaatagtcaagtcagcagccagccttctgttgaaagcgtac 2461
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341  aacagagaatagctaaatagtcaagtcagcagccttctgttgaaagcgtac 2400

Db   2461  agcctgtccaagttcacacagtatattaagagctctacattcttgggaggtttcggg 2520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2522  agcctgtccaagttcacacagtatattaagagctctacattcttgggaggtttcggg 2581

Qy   2582  agtctgcacagaagtgcacacactcattgccacgtgaagttcc 2641
           |||||||||||||||||||||||||||||||||||||||||||
Db   2582  agtctgcacagaagtgcacacactcattgccacgtgaagttcc 2641

Db   2521  agtctgcacagaagtgcacacacctcattgccagcagcaaagtgactcgcacacgtgaagttcc 2580
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2642  tgacggcgaattctgtcgtgaagcacacctagtagtgacgaggaggaggaatgcttca 2701

Qy   2702  ggtgtcaggaaagttcattgatgaggcagaactacattctccggagcagagtcagaagtac 2761
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581  tgacggcgaattctgtcgtgaagcacactagtgacgaggaggaggaatgcttca 2640

Db   2641  ggtgtcaggaaagttcattgatgaggcagaactacattctccggagcagagtcagaagtac 2700
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2762  tttctctttcctctggaagaatcctcttaaaaacgggcacacagttcctccacctcttaagg 2821

Qy   2821  tttctctttcctctggaagaatcctcttaaaaacgggcacacagttcctccacctcttaagg 2880
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701  tttctctttcctctggaagaatcctcttaaaaacgggcacacagttcctccacctcttaagg 2760

Db   2761  caaaatattttacatcacactctggaatctgccaagtctctccaactatgaaggccaatcg 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2882  tagagttgtcaggaagggagggaagtgttatcccaagcacaatcttccggaagctcaatggagc 2941

Qy   2942  acaagcagaaactcgagctttgtgggaaaatgttgtatccaagcaagcaagcatcatttt 3001
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821  tagagttgtcaggaagggagggaagtgttatccaagcaatcttccggaagctcaatgggagc 2880

Db   2881  acaagcagaaactcgagctttgtgggaaaatgttgtatccaagcaagcaatcatttt 2940
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3002  tatgccgagagatattttgccagaggccatagaatctgttcacacaagtgtcgttctgactg 3061

Qy   3062  gagtgctcactcaaacgctgggactgagtcagtcaatctaactgatctgcctcatgggaa 3121
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941  tatgccgagagatattttgccagaggccatagaatctgttcacacaagtgtcgttctgactg 3000

Db   3001  gagtgctcactcaaacgctgggactgagtcagtcaatctaactgatcgcgctggaa 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3122  gccgtgcatgtcgcactcgcgtggggggctgcgggaggagactgcgc 3181

Qy   3182  gctcctccactgctcttcagctgccctgggggatcaagctgaagcag 3241
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061  gccgtgcatgtcgacttcctgcggtgcgggggctgcgggaggagactgcgc 3120

Db   3121  gctcctccactgctcttgttcctgggggatcaagctgaagcag 3180
           ||||||||||||||||||||||||||||||||||||||||||||
Qy   3242  gaagacagagatataatactgacatataatcatttaagatgaagctcgag 3301

Qy   3302  acataaatatatgtttgtattatatatgacttaagagcccaacattaggtttttatgattcat 3361
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181  gaagacagagatataatactgacatataatcatttaagatgaatcagttctggagaa 3240

Db   3241  acataaatatatgtttgtattatatgacttaagagcccaacattaggtttttatgattcat 3300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3362  ttgccagtgttttaaatgtttttcacacaaaacgttcacggactcaactagaaaataaaatg 3421

Qy   3422  gttgtaaattaaagacccttgctatctctaaatatgtggatgttaaagatgttttgt 3481
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301  ttgccagtgttttaaatgtttttcacacaaaacgttacgggactcaactagaaaataaaatg 3360

Db   3361  gttgtaaattaaagacccttgctatctctaaatatgtggatgttaaagatgttttgt 3420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   3482  acttgattatttcttttatttattctttcttttatactctgtttctttattgacatctttgcccacat 3541

Qy   3542  tttaaataaatgtactttgaacttagaaaaaaaaaa 3580
           ||||||||||||||||||||||||||||||||||||
Db   3421  acttgattatttcttttattctttcttttatactctgtttctttattgacatcttctttatttgcccacat 3480

Db   3481  tttaaataaatgtactttgaacttagaaaaaaaaaaaaaa 3519
           ||||||||||||||||||||||||||||||||||||
```

DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 31086.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999; 99US-0122487.
XX
PA   (GENS) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PS   Claim 1; SEQ ID 31086; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC   5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC   used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC   in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors.
XX
SQ   Sequence 335 BP; 118 A; 59 C; 84 G; 71 T; 3 other;

Query Match   9.3%; Score 331.2; DB 21; Length 335;
Best Local Similarity   98.8%;
           Matches 331; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   655 cagclaacctcaalaagaaaalgcacgcaattgattgttccagaggccaaagggggggaaa 714
          |||||||||||||| ||| |||||||||||||||||||||||||||||| ||||||| |||
Db   1 cagctaacctcaataagaaatgcacgcaattgattgttccagagccaaagggggggagaaa 60

Qy   715 tacgaatgtgcttaaagcgagcaagtattaaaaattgtgactgactggttctcggat 774
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 tacgaatgtgcttaaagcgagcaagtattaaaaattgtgactgactggttctcggat 120

Qy   775 tgcggtatcagagagaaaaaccaaaaagaacagaagcaagacaagcaatttttatcctctctctattattat 834
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ngcgtatcagagagaaaaccaaaaagaacagaagcaagacaagcaatttttatcctctctctattattat 180

Qy   835 gaagaggaagaagaggaggaggaggagaggagggaggacaagaggttat 894
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 gaagaggaagaagaggaggaggaggagaggagggaggacaagaggttat 240

Qy   895 cagaatgaggtagtacagatagatgaggaagtcaaggccctgccagctctcagagaaggttcctcct 954
          |||||||||||||||||||||||||||||||||||:|||||||||||||||:|||||||||||||
Db   241 cagaatgaggtagtacagatagatgaggaagtcaagccctkvcagctctcagagaaggtctctcct 300

Qy   955 tcaggtgaccagcagcagtttcacctaaaatccaacac 989
          ||||||||||||||||||||||||||||||||||||||
Db   301 tcaggtgaccagcagtgtttcacttaaatccaaac 335

RESULT 3
AAH70661
ID   AAH70661 standard; cDNA; 690 BP.
XX
AC   AAH70661;
XX
DT   19-SEP-2001 (first entry)
XX
DE   Human cervical cancer marker nucleic acid 1935.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000; 2000US-0188315.
PR   12-MAY-2000; 2000US-0203791.
PR   09-JUN-2000; 2000US-0203971.
PR   09-JUN-2000; 2000US-0210060.
PR   21-JUL-2000; 2000US-0220114.
XX
PS   Claim 1; Page 415; 105lpp; English.
XX
PI   Schlegel R, Deeds J, Berger A, Zhao X;
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
DR   WPI; 2001-375006/39.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 690 BP; 218 A; 174 C; 172 G; 124 T; 2 other;

Query Match   5.7%; Score 204.6; DB 22; Length 690;
Best Local Similarity   95.9%;
           Matches 210; Conservative   0; Mismatches   9; Indels   0; Gaps   0;

Qy   1073 cggccgaagtcccacagttagctgcagcaaacgctgcctcagggaaagggacctg 1132
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23 cggccgaagtcccacagttagctgcagcaaacgctgcctcagggaaagggacctg 82

Qy   1133 ggttgattaacttgtgtgccaatgtcccacccgtccaggtaacattttgccccctgagg 1192
          |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   83 ggttgattaacttgtgtcaatgtcccacccgtccaggtaacattttgccccctgagg 142

Qy   1193 tccggggtaatttaatggctgctggacaaaaacctccaaagttctggaaagatcagaaatga 1252
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 tccgggggtaaatttaatggctgctcggacaaaaacctccaaagttctggaaagatcagaaatga 202

Qy   1253 tagctacctgagtccaactgtacggacactgaggata 1291
          |||||||||||||||||| ||| ||||||||||||
Db   203 tagctacctgagtccaactgtacggcacttggcgtaa 241

RESULT 4

ID   AAH71865
XX
AC   AAH71865 standard; cDNA; 504 BP.
XX
AC   AAH71865;
XX
DT   19-SEP-2001 (first entry)
XX
DE   Human cervical cancer marker nucleic acid 3139.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000;   2000US-0189315.
PR   12-MAY-2000;   2000US-0203791.
PR   09-JUN-2000;   2000US-0210600.
PR   21-JUL-2000;   2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 622; 105ipp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 504 BP; 147 A; 135 C; 121 G; 101 T; 0 other;

Query Match   5.2%;   Score 186.8;   DB 22;   Length 504;
Best Local Similarity   94.0%;
Matches   205;   Conservative   0;   Mismatches   12;   Indels   1;   Gaps   1;

Qy   1075 gccgaagtcccaacagttagctcacacaaaacgcaagcgctgcctcagggaaggagcctggg 1134
            | | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21   ggccgagtcccacagttagctgcagcaaaacgcaggcgcctcagggaaggagcctggg 80

Qy   1135 ttgattaacttgtgtcaatgtcccacccgtcccaggtaacatttgccccctgagtc 1194
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81   ttgattaacttgtgtcaatgtcccacccgtcccaggtaacatttgccccctgagtc 140

Qy   1195 cggggtaattaatggctgctggacaaa???ccaaagttc-tgaaagatcagaaatgat 1253
            ||||||||| |||||||||||||||||||| ||||||| |||||||||||||||||
Db   141   cggggtaatttaatggctgctggacaaaacctcaaagttcttgaaagatcagaaatgat 200

Qy   1254 agctacctggagtccagctgtacggacactgaggaata 1291
            |||||||||||||||||||||||||| || | |
Db   201   agctacctggagtccagctgtacggcactggcgtaa 238

RESULT 5

AAH68998
ID   AAH68998 standard; cDNA; 579 BP.
XX
AC   AAH68998;
XX
DT   19-SEP-2001 (first entry)
XX
DE   Human cervical cancer marker nucleic acid 272.
XX
KW   Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS   Homo sapiens.
XX
XX
PN   WO200142467-A2.
XX
PD   14-JUN-2001.
XX
PF   08-DEC-2000; 2000WO-US33312.
XX
PR   08-DEC-1999;   99US-0169681.
PR   21-DEC-1999;   99US-0171350.
PR   14-MAR-2000;   2000US-0189315.
PR   12-MAY-2000;   2000US-0203791.
PR   09-JUN-2000;   2000US-0210600.
PR   21-JUL-2000;   2000US-0220114.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR   WPI; 2001-375006/39.
XX
PT   New isolated nucleic acid for diagnosing and treating cervical cancer
PT   and for assessing and detecting compounds for treating the cancer -
XX
PS   Claim 1; Page 150-151; 105ipp; English.
XX
CC   The invention relates to novel genes (AAH68727-AAH73383) associated with
CC   cervical cancer with cytostatic activity. The nucleic acids and encoded
CC   polypeptides are useful: to assess if a patient is afflicted with
CC   cervical cancer or has a pre-malignant condition; to monitor the
CC   progression of cervical cancer or a premalignant condition in a patient;
CC   and to select and/or assess the efficacy of a compound or therapy for
CC   inhibiting cervical cancer in a patient. The nucleic acids may also be
CC   useful for gene therapy.
XX
SQ   Sequence 579 BP; 169 A; 152 C; 145 G; 110 T; 3 other;

Query Match   5.2%;   Score 186.8;   DB 22;   Length 579;
Best Local Similarity   94.0%;
Matches   205;   Conservative   0;   Mismatches   12;   Indels   1;   Gaps   1;

Qy   1075 gccgaagtcccaacagttagctgcagcaaaacgcaagcgctgcctcagggaaggagcctggg 1134
            | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21   ggccgagtcccacagttagctgcagcaaaacgcaggcgcctcagggaaggagcctggg 80

Qy   1135 ttgattaacttgtgtcaatgtcccacccgtcccaggtaacatttgccccctgagtc 1194
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81   ttgattaacttgtgtcaatgtcccacccgtcccaggtaacatttgccccctgagtc 140

Qy   1195 cggggtaattaatggctgctggacaaaacctccaaagttc-tgaaagatcagaaatgat 1253
            ||||||||| |||||||||||||||||||||||||||| |||||||||||||||||
Db   141   cggggtaatttaatggctgctggacaaaacctcaaagttcttgaaagatcagaaatgat 200

Qy   1254 agctacctggagtccagctgtacggacactgaggaata 1291
            |||||||||||||||||||||||||| || | |
Db   201   agctacctggagtccagctgtacggcactggcgtaa 238

RESULT 6
AAI85135
ID   AAI85135 standard; cDNA; 467 BP.

Left entry

```
XX
AC   AAI85135;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 5195.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO20016483S-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 5195; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI938841) and
CC   the encoded proteins (AAO00010-AAO13910), that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 467 BP; 122 A; 173 C; 107 G; 61 T; 4 other;

Query Match     4.5%;  Score 162;  DB 22;  Length 467;
Best Local Similarity   94.4%;
Matches  168;  Conservative   0;  Mismatches  10;  Indels   0;  Gaps   0;

Qy   1472 tacaagcagcagcagccagcagcagcagcaccggtttcacacctcagc 1531
          ||||||||||||| ||| |||||||||||||||||||||||||||||
Db     40 tacagcagcaggcgggccagcgagcagcagcagcacccggtttacacctcagc 99

Qy   1532 cccagcagagatataatgcagctccagcagcagcagcatctcagcaacctacc 1591
          ||||||||| ||||||||| ||||||||||||||||||||||||||||||||
Db    100 cccagcagatataatgcagctccagcagcagcagcgcagatctctcagcaacctacc 159

Qy   1592 cccagcagcccccgatccattcacgcagcagcagcagcagcagcagcaccacc 1649
          |||||||||| ||| |||||||||||||||||||||||||||||| |||| ||
Db    160 cccagcagcgcgcccatccatttcacgcagcaacagcagcagcagccaagcccatcc 217
```

Right entry

```
RESULT 7
AAX11992

ID   AAX11992 standard; DNA; 150 BP.
XX
AC   AAX11992;
XX
DT   30-MAR-1999 (first entry)
XX
DE   Human biallelic polymorphic DNA fragment WI-18169.
XX
KW   Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW   detection; phenotypic typing; characteristic; infection; hereditary;
KW   autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW   treatment; marker; ss.
XX
OS   Homo sapiens.
XX
PN   WO9820165-A2.
XX
PD   14-MAY-1998.
XX
PF   05-NOV-1997; 97WO-US20313.
XX
PR   06-NOV-1996; 96US-0030455.
XX
DR   WPI; 1998-286974/25.
XX
PI   Hudson T, Lander ES, Wang D;
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PT   New isolated nucleic acid segments from the human genome - used for
PT   determining polymorphic forms for use in e.g. forensics, paternity
PT   testing or phenotypic typing for disease
XX
PS   Claim 1; Page 210; 310pp; English.
XX
CC   AAX10269-X12937 are human DNA fragments which contain biallelic
CC   polymorphic markers which have been isolated using the primers
CC   represented in AAX09121-X10268. The base occupying the polymorphic site
CC   is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC   can be used in methods for determining polymorphic forms in an individual
CC   for use in e.g. forensics, paternity testing or for phenotypic typing for
CC   diseases such as agammaglobulinaemia, diabetes insipidus, Lesch-Nyhan
CC   syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC   familial hypercholesterolemia, polycystic kidney disease, hereditary
CC   spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC   haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC   syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC   autoimmune diseases, inflammation, cancer, diseases of the nervous
CC   system, infection by pathogenic microorganisms, and characteristics such
CC   as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC   endurance, fertility, and susceptibility or receptivity to particular
CC   drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC   segments can also be used to produce medicaments for the treatment or
CC   prophylaxis of such diseases.
XX
SQ   Sequence 150 BP; 43 A; 32 C; 40 G; 34 T; 1 other;

Query Match     3.8%;  Score 135.6;  DB 19;  Length 150;
Best Local Similarity   99.3%;
Matches  135;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   2820 ggcaaaatattttacatcacacctgccaagtctttccactatgaaggcaat 2879
          | |||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ggcaaaatattttacatcacacctggaatctgccaagtctttccactatgaaggcaat 60

Qy   2880 cgtagagtgtgcagagagaaggtgttatccaagcagcacatctccgaaagtcatgga 2939
          ||||||||||||||| ||||||||||||||||||||||||||| |||||| ||||||
Db     61 cgtagagtgtgcagaggaggaaggtgttatccaagcagcacatctttccggaaagctcctgga 120

Qy   2940 gcacaagcagaactcg 2955
          ||||||||||||||||
Db    121 gcacaagcagaactcg 136
```

RESULT 8
AAS44894/c
XX
ID   AAS44894 standard; DNA; 344 BP.
XX
AC   AAS44894;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human contig polynucleotide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation; dsi
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-598862/66.
DR   P-PSDB; AAU27994.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection.
XX
PS   Claim 1; SEQ ID No 491; 153pp; English.
XX
CC   Sequences AAS44576-AAS44919 represent full-length polynucleotides and
CC   contig polynucleotides encoding polypeptides of the invention. The DNA
CC   and protein sequences are useful for the treatment, diagnosis and
CC   prevention of various types of disorder in a mammalian subject such as a
CC   human, dog, monkey, mouse, hamster or rat. The disorders include cancers
CC   such as leukaemia, lymphoma and neuroblastoma, autoimmune disorders such
CC   as multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.

CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 344 BP; 97 A; 83 C; 79 G; 85 T; 0 other;

Query Match   3.5%; Score 125; DB 22; length 344;
Best Local Similarity   100.0%; Pred. No. 4e-19;
Matches 125; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   2831 tttacatcacacctggaatctgcccaagtcttccacctatgaaggccaatcgtgtaagtagtgtg 2890
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 TTTACATCACACCTGGAATCTGCCCAAGTCTTCCACCTATGAAGGCCAATCGTGTAAGTAGTGTG 285

Qy   2891 caggaggaaaggtgttatccaagcagccaatctcttccggaagctcaatggagcacaagcaga 2950
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 CAGGAGGAAAGGTGTTATCCAAGCAGCCAATCTCTTCCGGAAGCTCAATGGAGCACAAGCAGA 225

Qy   2951 actcg 2955
            |||||
Db   224 ACTCG 220

RESULT 9
AAX89891
XX
ID   AAX89891 standard; DNA; 397 BP.
XX
AC   AAX89891;
XX
DT   05-NOV-1999 (first entry)
XX
DE   Spinocerebellar ataxia type III (SCAIII) gene fragment.
XX
KW   Spinocerebellar ataxia type III; SCAIII; reverse dot hybridisation;
KW   PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
KW   SCAIII syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FH
FT   repeat_region   137..355
FT   /*tag= a
FT   repeat_unit   137..139
FT   /*tag= a
FT   /*tag= b
FT   /note= "trinucleotide repeat"
XX
PN   WO9943852-A1.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999; 99WO-KR00078
XX
PR   26-FEB-1998; 98KR-0006278.
XX
PA   (JIND/) JIN D K.
PA   (SMSU) SAMSUNG FINE CHEM CO LTD.
XX
PI   Jin DK;
XX
DR   WPI; 1999-527634/44.
XX
PT   Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
PT   using techniques which ensure highly accurate diagnosis
XX
PS   Claim 1; Page 12-13; 28pp; English.
XX
CC   The invention relates to the diagnosis of spinocerebellar ataxia type
CC   III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
CC   hybridisation (PCR-MPH). The method comprises attaching a potion of the
CC   SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
CC   (the present sequence) to a substrate, and hybridising with amplified
CC   testee genomic DNA containing copies of the trinucleotide units, PCR

CC   amplified with labeled primers (AAX89889-90). The new method is useful
CC   for diagnosis of SCAIII syndrome, and for determining the severity of
CC   the disease. The present sequence represents the SCAII gene fragment
CC   containing 73 trinucleotide (TNR) repeats.
XX
SQ   Sequence 397 BP; 124 A; 104 C; 104 G; 65 T; 0 other;

Query Match   3.1%;   Score 110.6;   DB 20;   Length 397;
Best local Similarity   63.28;   Pred. No. 9.3e-16;
Matches 170; Conservative   0; Mismatches 99; Indels 0; Gaps 0;

Qy 1435 tttagccaagtgaaagtgactccagagaacaacacatgctacagc.agcagcaggcccag 1494
Db 113 ttcagacaagcaaagtgaaagtgactccagagaacaacacatgctacagcagcagcagcag 172

Qy 1495 cagcagcagcagcagcaccggttttacacctttagcccagcagcagcagcagcttccag 1554
Db 173 cagcagcagcagcaccggttttacacctttagcccagcagcagcagcagcttccag 232

Qy 1555 cagcagcagcagcagcagcaatctcttcagcaacttacccccagcagcagcagcatcatt 1614
Db 233 cagcagcagcagcagcagcaatctcttcagcaacttacccccagcagcagcatcatt 292

Qy 1615 tcacacagcagcagcaacaacagcaccaccacatcgcctcaacagcagcaatcagcttttt 1674
Db 293 cagcagcagcagcagcaacaacagcaccaccacatcgcctcaacagcagcaatcagc 352

Qy 1675 ggacatgatccagccagtgggagattccaga 1703
Db 353 caacgggacttatcaggacaagagttcaca 381

RESULT 10
AAC84677
ID AAC84677 standard; DNA; 5120 BP.
XX
AC AAC84677;
XX
DT 20-APR-2001 (first entry)
XX
DE DNA sequence of human PPP2R2B.
XX
KW Microsatellite marker; human; chromosome 5q31-33; D5S402; WI-6763;
KW CAG trinucleotide repeat; spinocerebellar ataxia 12; SCA12; PPP2RBbeta;
KW brain; protein phosphatase PP2A; ds.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FH
FT repeat_region 2008..2366
FT   /*tag= a
FT   /rpt_type= "CAG"
FT mRNA   2473..3093
FT   /*tag= b
FT 5'UTR   2473..3023
FT   /*tag= c
FT CDS   3024..3093
FT   /*tag= d
FT   /product= "PPP2R2B"
FT   /note= "protein_id= AAF74024.1"
XX
PN WO20007894J-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US40213.
XX
PR 18-JUN-1999; 99US-0140176.
XX
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Margolis RL, Ross CA, Holmes SE;

XX
DR WPI; 2001-102718/11.
DR P-PSDB; AAB48340.
XX
PT New polynucleotide for use in diagnosing spinocerebellar ataxia 12,
PT comprises a microsatellite marker having a variable number of CAG
PT trinucleotide repeats located on human chromosome 5q31-33 -
XX
PT Disclosure; Page 21-23; 23pp; English.
XX
CC The invention relates to a microsatellite marker located on human
CC chromosome 5q31-33 between markers D5S402 and WI-6763, comprising a
CC variable number of CAG trinucleotide repeats. The marker is useful for
CC diagnosing spinocerebellar ataxia 12 (SCA12) by determining the number of
CC CAG trinucleotide repeats in a 5'-untranslated region of an allele of
CC PPP2RBbeta, a brain specific regulatory subunit of protein phosphatase
CC PP2A, where if a number greater than 40 is determined, an expansion which
CC is associated with SCA12 is indicated. The presence of expanded
CC trinucleotide repeats on chromosome 5q31-33 is useful for predicting or
CC diagnosing SCA12. Also provided are primers useful for diagnosing SCA12 and
CC microsatellite marker associated with SCA12 and for amplifying the
CC microsatellite marker located on human chromosome 5q31-33. The present
CC a probe useful for detecting expansions in a microsatellite marker
CC located on human chromosome 5q31-33. The present sequence represents
CC a PPP2R2B DNA sequence.
XX
SQ Sequence 5120 BP; 1336 A; 1267 C; 1268 G; 1249 T; 0 other;

Query Match   2.9%;   Score 105.6;   DB 22;   Length 5120;
Best Local Similarity   62.5%;   Pred. No. 4.2e-14;
Matches 165; Conservative   0; Mismatches 99; Indels 0; Gaps 0;

Qy 1406 agggacatacaaatgccaatgcagtgtctgtttagccaagtgaactgcaactccagagacac 1465
Db 2017 agagtcgtg ggctgctgacgcggttgggaggaggagcctcgcctttaatgcacccgcct 2076

Qy 1466 acatgctacagcagcagcaggcccagcagcagcagcaccgtttacacc 1525
Db 2077 ccagcctcctgcagcacagcagcagcagcagcagcagcagcagcagcagc 2136

Qy 1526 ttcagcccagcagataatgcagtccagcagcagcagcagatctctcagcaac 1585
Db 2137 agcagcagcagcagccccgcagcagcagcagcagcagcagcagc 2196

Qy 1586 cttacccccagcagccccgcgatccatttcacagcagcagcagcagcac 1645
Db 2197 agcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcac 2256

Qy 1646 cacatcccctcagcagcagtcagc 1669
Db 2257 agcagcagcagcagcagcagcagc 2280

RESULT 11
AAA59242/c
ID AAA59242 standard; cDNA; 1037 BP.
XX
AC AAA59242;
XX
DT 07-NOV-2000 (first entry)
XX
DE Exons E, C and A of the SCA8 repeat region.
XX
KW Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS Homo sapiens.
XX
PN CA2283758-A1.
XX
PD 28-APR-2000.
XX
PF 27-SEP-1999; 99CA-2283758.

Left column

```
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
DR   WPI; 2000-491456/44.
XX
PI   Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
WPI: 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
DR   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 25; Fig 7C; 68pp; English.
XX
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1037 BP; 229 A; 261 C; 242 G; 305 T; 0 other;
```

```
Query Match    2.9%;    Score 105.2;  DB 21;   Length 1037;
Best Local Similarity   69.4%;   Pred. No. 2.5e-14;
Matches   143;  Conservative   0;  Mismatches   63;  Indels   0;  Gaps   0;

Qy   1464 acacatgctacagcagcagcagcccagcagcagcagcagcagcacccggtttaca 1523
           |||| | |||||||||||||||||||||||||||||||||||||| ||||||
Db   920  AAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 861

Qy   1524 ccttcagcgcccagcagagtaatgcagctccagcagcagcagcagatctctcagca 1583
           |||| |||||||||||||||||||||||||||||||||||| || |||||
Db   860  GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 801

Qy   1584 accttacccccagcagcgcccgcatccatttcacagcagcagcagcaacagcagcc 1643
           |||| | |||||||| |||||||||||||||||| || || |||||
Db   800  CCAGCAGCAGCAGCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 741

Qy   1644 accaccatcgcctcaagcagcatcagc 1669
           ||| |||||||||| ||||
Db   740  GCAGCAGCAGCAGCAGCAGCAGCAGCAGC 715
```

```
RESULT 12
ID   AAA59240/c
AC   AAA59240;
XX
AAA59240 standard; DNA; 1159 BP.
XX
DT   07-NOV-2000   (first entry)
XX
DE   An EcoRI fragment that includes SCA8 repeat region.
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
```

Right column

```
XX
PI   Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR   WPI; 2000-491456/44.
XX
PT   Novel nucleic acid molecule containing repeat region of spinocerebellar
PT   ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT   or determining risk for developing spinocerebellar ataxia -
XX
PS   Claim 26; Fig 7A; 68pp; English.
XX
CC   The present sequence represents a nucleic acid molecule containing a
CC   repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC   sequence which is located within the long arm of chromosome 13. The
CC   SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC   Generally, the CTA repeat is unstable, and can vary between individuals
CC   in different families (the number of CTA repeats in the region does not
CC   vary between individuals of a family). Fragments of the nucleic acid
CC   molecule are useful as primers and probes for determining whether an
CC   individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ   Sequence 1159 BP; 263 A; 263 C; 196 G; 437 T; 0 other;
```

```
Query Match    2.9%;    Score 105.2;  DB 21;   Length 1159;
Best Local Similarity   69.4%;   Pred. No. 2.7e-14;
Matches   143;  Conservative   0;  Mismatches   63;  Indels   0;  Gaps   0;

Qy   1464 acacatgctacagcagcagcagcccagcagcagcagcagcagcacccggtttaca 1523
           |||| | |||||||||||||||||||||||||||||||||||||| ||||||
Db   732  AAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 673

Qy   1524 ccttcagcgcccagcagagtaatgcagctccagcagcagcagcagatctctcagca 1583
           |||| |||||||||||||||||||||||||||||||||||| || |||||
Db   672  GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGGAGCAGCA 613

Qy   1584 accttacccccagcagcgcccgcatccatttcacagcagcagcagcaacagcagcc 1643
           |||| | |||||||| |||||||||||||||||| || || |||||
Db   612  GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 553

Qy   1644 accaccatcgcctcaagcagcatcagc 1669
           ||| |||||||||| ||||
Db   552  GCAGCAGCAGCAGCAGCAGCAGCAGCAGC 527
```

```
RESULT 13
ID   AAA59241/c
AC   AAA59241;
XX
AAA59241 standard; cDNA; 1472 BP.
XX
DT   07-NOV-2000   (first entry)
XX
DE   Exons D, C, B and A of the SCA8 repeat region.
KW   Repeat region; spinocerebellar ataxia type 8; SCA8; long arm;
KW   chromosome 13; polymorphic CTA repeat; CTG repeat; ss.
XX
OS   Homo sapiens.
XX
PN   CA2283758-A1.
XX
PD   28-APR-2000.
XX
PF   27-SEP-1999;   99CA-2283758.
XX
PR   28-OCT-1998;   98US-0181585.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Benzow KA, Moseley-alldredge ML, Ranum LPW, Koob MD;
XX
DR   WPI; 2000-491456/44.
```

XX
PT Novel nucleic acid molecule containing repeat region of spinocerebellar
PT ataxia type 8 coding sequence, useful as primer or probe for diagnosing
PT or determining risk for developing spinocerebellar ataxia -
XX
PS Claim 24; Fig 7B; 68pp; English.
XX
CC The present sequence represents a nucleic acid molecule containing a
CC repeat region of a spinocerebellar ataxia type 8 (SCA8) coding
CC sequence which is located within the long arm of chromosome 13. The
CC SCA8 coding sequence contains polymorphic CTA repeats and CTG repeats.
CC Generally, the CTA repeat is unstable, and can vary between individuals
CC in different families (the number of CTA repeats in the region does not
CC vary between individuals of a family). Fragments of the nucleic acid
CC molecule are useful as primers and probes for determining whether an
CC individual has or is at risk of developing spinocerebellar ataxia 8.
XX
SQ Sequence 1472 BP; 337 A; 368 C; 378 G; 388 T; 1 other;

Query Match   2.98;   Score 105.2; DB 21; Length 1472;
Best Local Similarity   69.4%; Pred. No. 3e-14;
Matches 143; Conservative   0; Mismatches 63; Indels   0; Gaps   0;

```
Qy  1464 acacatgctacagcagcagcaggcagcagcagcccagcagcagccaccccggtttaca 1523
          || ||  | |||||||||| |||||||||  |||||||||| ||
Db  1355 AAAAAATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1296

Qy  1524 ccttcagcccccagcagaatatgcgagtccagcagcagcagcagcagatctctcagca 1583
          |||||| ||||||||
Db  1295 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 1236

Qy  1584 accttacccccagcagccgccgccatccatttcacacagcagcagcaaacagcagcc 1643
Db  1235 GCAGCAGCAGCACCAGCACCACCACCACCACCACCAGCAGCAGCAGCAGCAGCAGCA 1176

Qy  1644 accaccatcagcctcaagcagcaatcagc 1669
Db  1175 GCAGCAGCAGCACCAGCAGCAGCAGCAGCAGCAC 1150
```

RESULT 14
ID AAQ84832 standard; DNA; 234 BP.
XX
AC AAQ84832;
XX
DT 25-SEP-1995 (first entry)
XX
DE Spinocerebellar ataxia type 1 CAG repeat region patient #2.
XX
KW Spinocerebellar ataxia type 1; SCA 1; presymptomatic diagnosis;
KW CAG repeat region; patient #2; ss.
XX
OS Homo sapiens.
XX
PN WO9501437-A.
XX
PD 12-JAN-1995.
XX
PF 29-JUN-1994;   94WO-US07336.
PR 29-JUN-1993;   93US-0084365.
PR 28-JUN-1994;   94US-0267803.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Chung M, Orr HT, Zoghbi HY;
XX
DR WPI; 1995-061001/08.
XX
PT New autosomal dominant spinocerebellar ataxia type 1 nucleic acid
PT - used to develop prods. for detection or presymptomatic

PT diagnosis of a SCA1 disorder
XX
PS Disclosure; Fig 2; 111pp; English.
XX
CC AAQ84831-Q84835 show the CAG repeat regions of five individuals
CC affected with spinocerebellar ataxia type 1 (SCA 1). It is within
CC SCA 1 ... The full nucleic acid (AAQ84793) and its protein product
CC CAG repeat region (AAQ84804) that the mutations responsible for
CC (AAP71111) can be used to develop products, for the presymptomatic
CC detection of a SCA 1 disorder.
XX
SQ Sequence 234 BP; 73 A; 80 C; 77 G; 4 T; 0 other;

Query Match   2.98;   Score 103; DB 16; Length 234;
Best Local Similarity   66.4%; Pred. No. 4.2e-14;
Matches 148; Conservative   0; Mismatches 75; Indels   0; Gaps   0;

```
Qy  1465 cacatgctacagcagcagcagcagcccagcagcagcagcagcagcagcaccccggtttacac 1524
Db     4 cagcagcagcagcagccgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 63

Qy  1525 cttcagcccccagcagaataatgcgagtccagcagcagcagcagcagatctctcagcaa 1584
Db    64 cagcagcagcagcagccgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 123

Qy  1585 ccttacccccagcagccgccgccatccatttcacacagcagcagcagcaaacagcagcca 1644
Db   124 cagcagcagcagcagccgcagcagcagcagcagcagcagcagcagcagcagcagcagcag 183

Qy  1645 ccaccatcagcctcagcagcaatcagctttttggacatgatccag 1687
Db   184 cagcagcagcagcagcagcagcagcacctcagcaggctccgg 226
```

RESULT 15
ID AAZ44307 standard; DNA; 477 BP.
XX
AC AAZ44307;
XX
DT 04-APR-2000 (first entry)
XX
DE Human SCA7 genomic DNA fragment 1.
XX
KW SCA7; human; spinocerebellar ataxia type 7; SCA1; SCA2; SCA3; SCA6;
KW repeat expansion detection; RED analysis; detection; ds.
XX
OS Homo sapiens.
XX
PN CA2245310-A.
XX
PD 19-FEB-1999.
XX
PF 19-AUG-1998;   98CA-2245310.
PR 19-AUG-1997;   97US-0056170.
XX
PA (MINU) UNIV MINNESOTA.
XX
PI Koob MD, Ranum LP;
XX
DR WPI; 2000-098181/09.
XX
PT Identifying individuals at risk of developing spinocerebellar ataxia
PT type 7 by analyzing trinucleotide repeat regions of spinocerebellar
PT ataxia type 7 gene -
XX
PS Claim 36; Page 41; 66pp; English.
XX
CC This invention describes a novel method for identifying individuals at
CC risk for developing spinocerebellar ataxia type 7 (SCA7). The method
CC comprises analyzing the CAG repeat region of a SCA7 gene to detect CAG

CC repeats, where individuals at risk have at least 30 CAG repeats and
CC those not at risk have less than 19 CAG repeats. The method is useful
CC for identifying individuals at risk of developing SCA7 and also those
CC at risk of developing SCA1, 2, 3 or 6. The use of genomic DNA in the
CC repeat expansion detection (RED) analysis allows isolation of any
CC potential trinucleotide repeat expansion regardless of the expression
CC pattern. Utilization of different oligonucleotides in the RED assay
CC allows any of the possible trinucleotide repeats to be detected, and
CC the cycled nature of the reaction makes it extremely sensitive. This
CC sequence represents a genomic DNA fragment of the human SCA7 gene
CC which is described in the method of the invention.
XX
SQ Sequence 477 BP; 99 A; 164 C; 172 G; 42 T; 0 other;

Query Match   2.8%; Score 100.2; DB 21; Length 477;
Best Local Similarity 69.9%;
Matches 135; Conservative   0; Mismatches 58; Indels   0; Gaps   0;
                             Pred. No. 2.6e-13;

Qy 1477 cagcagcagcagccagcagcagcagcagcaccggtttacacttcagccccag 1536
         | |||||||||| |||||||||||||||||||||| | | |||| |||
Db   124 cggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 183

Qy 1537 cagataatgcagtccagcagcagcagcagatctcagcaaacttacccccag 1596
         ||| |||||| |||||||||||||||||||||||||||| || |||
Db   184 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 243

Qy 1597 cagcgccgcatccatttcacagcagcagcagcaacagcagccaccaccatccgcct 1656
         ||||||||| ||| |||||||||||||||||||||||||||||| |||| | |||
Db   244 cagcagcagcagcagcagcagcagcagcagcagcagcccgccgcctccgcagccc 303

Qy 1657 cagcagcatcagc 1669
         ||| ||| |||
Db   304 cagcggcagcagc 316

Search completed: June 4, 2002, 15:45:30
Job time: 10778 sec

OM of: US-09-664-641-10 to: SwissProt_40:*   out_format : pfs

Date: Jun 4, 2002 5:50 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110933_9290/app_query.fasta_1.3697
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsp -CAPOP=12.000
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP-4.500 -QGAPEXT-0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09664641 @CGN1_1_66 -NCPU=6 -ICPU-3 -LONGLOG
-DEV_TIMEOUT-120 -WARN-TIMEOUT-30 -NO_XLPXV -WAIT -THREADS=1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: SwissProt_40:*
Database sequences: 105224
Database length: 38719550
Search time (sec): 74.270000

score_list:

Sequence	Strd	Orig	ZScore	FScore	Len	! Documentation
SwissProt_40:YBM5_SCHPO	+	314.50	334.75	6.8e-12	878	! Q10337 schizosaccharomyces
SwissProt_40:SNF5_YEAST	+	223.50	234.90	2.4e-06	905	! P18480 saccharomyces cerevis
SwissProt_40:YKF4_YEAST	+	213.00	225.17	1.0e-05	738	! P35732 saccharomyces cerevis
SwissProt_40:YO83_CAEEL	+	204.50	213.24	3.5e-05	1000	! P34619 caenorhabditis elega
SwissProt_40:YI92_HUMAN	+	200.00	201.82	7.1e-05	2124	! Q93074 homo sapiens (human)
SwissProt_40:FXP2_MOUSE	+	198.50	209.59	7.8e-05	714	! P58463 mus musculus (mouse).
SwissProt_40:FXP2_HUMAN	+	196.00	206.84	0.0001	715	! O15409 homo sapiens (human).
SwissProt_40:MCM1_YEAST	+	192.00	210.37	0.0002	286	! P11746 saccharomyces cerevis
SwissProt_40:KNQ1_YEAST	+	189.50	199.24	0.0003	756	! P53894 saccharomyces cerevis
SwissProt_40:SSN6_YEAST	+	186.00	193.30	0.0005	966	! P14922 saccharomyces cerevis
SwissProt_40:GLTA_WHEAT	+	185.50	201.37	0.0005	356	! P10385 tritium aestivum (wh
SwissProt_40:BRD4_HUMAN	+	182.50	186.50	0.0008	1362	! O60885 homo sapiens (human)
SwissProt_40:INVO_PIG	+	182.00	197.76	0.0007	347	! P18175 sus scrofa (pig). invol
SwissProt_40:KAPC_DICDI	+	181.50	191.82	0.0008	648	! P34099 dictyostelium discoi
SwissProt_40:GALY_KLULA	+	180.50	186.91	0.0010	1008	! P32257 kluyveromyces lactis
SwissProt_40:ICPO_HSVBJ	+	180.00	189.81	0.0010	676	! P29108 bovine herpesvirus ty
SwissProt_40:GBF_DICDI	+	180.00	189.41	0.0010	708	! P36417 dictyostelium discoide
SwissProt_40:OVO_DROME	+	178.00	184.00	0.0014	1028	! P51521 drosophila melanogast
SwissProt_40:IXR1_YEAST	+	177.50	188.15	0.0015	597	! P33417 saccharomyces cerevis
SwissProt_40:KRUB_HUMAN	+	176.50	196.76	0.0015	194	! O75690 homo sapiens (human).
SwissProt_40:KCN3_HUMAN	-	176.50	185.25	0.0017	736	! Q9ugi6 homo sapiens (human).
SwissProt_40:YM38_YEAST	+	176.50	184.99	0.0017	758	! Q03825 saccharomyces cerevis
SwissProt_40:CI21_HUMAN	+	176.00	182.98	0.0019	898	! Q9ulv3 homo sapiens (human).
SwissProt_40:ICPO_HSVBK	+	175.00	184.34	0.0021	676	! P29836 bovine herpesvirus ty
SwissProt_40:GALY_YEAST	+	174.50	179.74	0.0024	1081	! P29859 saccharomyces cerevi
SwissProt_40:YSH_DROME	+	173.00	172.62	0.0031	2038	! P13709 drosophila melanogast
SwissProt_40:ABF1_HUMAN	+	172.50	166.92	0.0036	3703	! Q15911 homo sapiens (human)
SwissProt_40:E78A_DROME	+	172.00	178.94	0.0033	864	! P45447 drosophila melanogast
SwissProt_40:KRUA_HUMAN	-	171.50	192.48	0.0030	169	! P26371 homo sapiens (human)
SwissProt_40:VIT_ICHUN	+	171.50	171.94	0.0038	1823	! Q91062 ichthyomyzon unicuspu
SwissProt_40:VIT1_CHICK	+	171.50	171.53	0.0038	1912	! P87498 gallus gallus (chick
SwissProt_40:SWI1_YEAST	+	170.50	173.68	0.0043	1314	! P09547 saccharomyces cerevi
SwissProt_40:MAM_DROME	+	170.50	172.00	0.0043	1596	! Q24742 drosophila melanogast
SwissProt_40:TRX_DROVI	+	170.50	164.45	0.0048	3828	! Q24742 drosophila melanogast
SwissProt_40:NAB2_YEAST	+	169.00	179.96	0.0048	525	! P21519 drosophila virilis (t
SwissProt_40:ATX1_HUMAN	+	169.00	176.15	0.0050	816	! P32505 saccharomyces cerevi
SwissProt_40:PRO_DROVI	+	168.50	170.03	0.0057	1556	! P54253 homo sapiens (human).
SwissProt_40:PHP_DROME	+	168.50	169.85	0.0057	1589	! Q9u6a1 drosophila virilis (t
SwissProt_40:TCOF_HUMAN	+	168.00	170.33	0.0061	1411	! P39769 drosophila melanogast
SwissProt_40:SIMA_DROME	+	168.00	169.76	0.0061	1507	! Q13428 homo sapiens (human)
						! Q24167 drosophila melanoga

SwissProt_40:YQC1_CAEEL	+	167.50	179.31	0.0058	468	! Q09260 caenorhabditis ele
SwissProt_40:RBF1_CANAL	+	167.50	178.28	0.0059	527	! Q00312 candida albicans (
SwissProt_40:NFT5_HUMAN	+	167.50	169.07	0.0066	1531	! O94916 homo sapiens (hum
SwissProt_40:ABF1_MOUSE	+	167.00	160.85	0.0078	3726	! Q94916 mus musculus (mou
SwissProt_40:YFAE_SCHPO	+	166.00	176.76	0.0072	520	! O14079 schizosaccharomyce

seq_name: SwissProt_40:YBM5_SCHPO

seq_documentation_block:

ID   YBM5_SCHPO   STANDARD;   PRT;   878 AA.
AC   Q10337;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 98.4 kDa protein C582.05c in chromosome II.
GN   SPBC582.05C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-972;
RA   Wood V., Rajandream M.A., Barrell B.G., Davis P., Churcher C.M.;
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
-!- SIMILARITY: SOME, TO YEAST YHR154W.
-!- SIMILARITY: CONTAINS 5 BRCT DOMAINS.

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an e-mail to license@isb-sib.ch).

DR   EMBL; AL096788; CAB46668.1; -.
DR   InterPro; IPR001357; BRCT.
DR   Pfam; PF00533; BRCT; 6.
DR   SMART; SM00292; BRCT; 5.
DR   PROSITE; PS50172; BRCT; 5.
KW   Hypothetical protein; Repeat.
FT   DOMAIN   14   104   BRCT 1.
FT   DOMAIN   128   196   BRCT 2.
FT   DOMAIN   311   399   BRCT 3.
FT   DOMAIN   695   744   BRCT 4.
FT   DOMAIN   767   864   BRCT 5.
SQ   SEQUENCE   878 AA; 98435 MW; AFCAFBEF800CBFC6 CRC64;

alignment_block:
US-09-664-641-10 to: SwissProt_40:YBM5_SCHPO

Align seq 1/1 to: YBM5_SCHPO   from: 1   to: 878

alignment_scores:
                Quality: 314.50   Length: 1194
Percent Similarity: Ratio: 0.758   Gaps: 35
Percent Similarity: 34.757   Percent Identity: 17.169

144 TACGGATCTCTCGGCCCCCTGCCCTTCAAATCACCTCTAGGCACAGGTTAT 193
     |||::::::|||| ||| :::||||||::|||| | ::
25   TyrSerAlaLeuGlnPro...AsnGlnAsnGluLeuArgLysLysGlue 40

194 TCAGCTTCTCAAGGCTGGAAAAGCGAAGGAAGGAGTTCCTACAAT...GCAC 240
     |:::|||||||||| ||||||||| ::: ||| |||: |
40   uPheIleLysAsnAspGlyLysAlaLeuSerPheProTyrAspTrpLysL 57

241 TAGCCTCACATAAATCTCAGAGATGGGACGAAATCCAGAGGTGGGAGAA 290
     ||||| ||| | |||||||||||||| || ::: |||||||| :
57   euAlaThrIleValIleCysAspAspPheSerSerProAsnValGlnGlu 73

```
 291 GCTCCGGAAGTCTTTGACTTACCTGTTGTTGTAAAGCCTTCTTCTTGGGTGATTCT  340
     :::  |||              :::::||||
  74 GlySerLys..ArgSerLeuArgLeuAlaLysThrAsnTrpIleArgAs          89

 341 GTCCGTTCAGTGTGGGAACTCTTCTGCCAGTAAATGGTTTTTCTCCAGAAT       390
     |||::|| |||          :|:||||
  89 pCysValAspLysAsnThrLeuLeuAsnTyrSerPheTyrSerCysAsnP        106

 391 CAGTCAGATTTTTTTTGAATACACTGCCTGCCTTTCTCAG.GTGTTGAT         440
     |::|||   |||||||||||              |||||||||
 106 roTyrLeuLeuPheLysGlyIleCysAlaSerSerCysGln..IleAsp         121

 441 ACAACCTGGAGCTCTTCGTTGGACTCTTCCAGAGCCTTCCCAAGGAGGAGG       490
     :::       |||||||||||||
 122 SerTyrGlnSerSerLeuIleAsp...........                       129

 491 TAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGGAAGCACAGAGATCATTCTG       540
 129 ..................................................       129

 541 CCTTCTTCTGACCCGGTATTGAATGACAGGGTGGAGGCCCTCTGTGTAATGTG      590
     |||||||||          |||::|||||
 130 ..................Asp.....Ala.LeuGluThr...........        134

 591 CTGGGTGTCATCTGAAGACAGAAGTGCCCCTGTGGGCTTTGGTTACGTTCT       640
 135 ..................................................p       135

 641 ATGGGGGAGAATTGCCAGCTAACCCTCAATAAGAAAATGCACGCATTTGATT      690
     |:||||||                     |||
 135 ...................................................       135

 691 GTTCAGAGCCAAAAGGGGGGAAATACGACATGTGCTTTAAAGCGA           735
     :::           |||||||::|||
 152 ThrTyrSerGlyMetGlyAlaLysCysLysValLeuAspLysProSe           168

 736 .GCAAGTAATTAAAATTGTGACTCCTGACTTCTGGATTCGGTATCAG          784
     ||||||||||               |||||||||::::|
 168 rLeuSerIleLysLeuIleHisProGlnThrpLeuLeuAspCysLeuGlnP      185

 785 AGAAAACCAAAAAGGACGAAGCAATTTTATCATCCTCGTCGTGATTATTTAT      834
     ||||||||||
 185 heGlyGlnLeuIleAspGln............                         191

 835 GAAGGAGAAGAAGNGGAAGNGAAGNGGAGNGGGAAGTAGAAAATGAGGA         884
 191 ...................................................      191

 885 ACAAGATTCTCAGAATGAGGGTAGTACGAGATGCAGAAGCTCAAGCGCCTGCCA    934
 191 ...................................................      191

 935 GCTCTCAGAGAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAATCC       984
     |||                      |||  |||||
 192 ...................AspProTyrLeuPheProAsnPro              199

 985 AACACTGAAAAATCTAAAGGGGAATTAAATGTTTGATGATTCTTCAGATTC       1034
     :::  ||||||        ::|||||
 200 SerTyrLysLys..............................AsnAspSe        206

1035 ATCACCGGAAAAACACGAGAGAAATTTAAACTGGACCCGCGGCGAAGTCC       1084
     |||||     |||||
 206 rSerIleSerLysAlaGlu.........................              212

1085 CACAGTTAGCTGCAGCAAAACGCAGCGGCTGCCTCAGGGAAAGAGGCCTGGG      1134
 212 ..................................................       212

1135 TTGATTAACTTGTGTGTGCCAAATGTCCCACCCGTCCCAGTAACATTTTGCC      1184
     |||||||||||
 213 ...........................ProThrSerLeu.....             216

1185 CCCTGAGGTCCGGGGTAATTAATTGGCTGCTGGACAAAACCTCCAAAGTT        1234
```

```
 216 ..................................................       216

1235 CTGAAAGATCAGAAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTG       1284
                                              |||
 217 ......................................Phe               217

1285 AGGAATAATTACTAATAATGCTGACATT...........CAGCAGATGAACCG    1325
     |||||||||||      :::::||||||      ::::::|||
 218 ArgAsnValLeuHisGlyLysArgIleTyrPheSerAsnAspLeuAsnLe       234

1326 GCCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAA       1375
     |||||||||||       |||  |||     ::|||
 234 uProThrAsnPheArgHisSerLeuGlnLys...........              244

1376 ATTTAGAACAGCAGGTGAATCACACGCCAGCACAGGGACCATACAAATGCCAAT    1425
 244 ..................................................       244

1426 GCAGTCGCTGTTTAGGCAAGTGAAAACTGACTCCAAGAACATGCTACA          1475
 244 ..................................................       244

1476 GCAGCAGCACAGGCCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACC          1525
 244 ..................................................       244

1526 TTCAGCCCCAGCAGATAAATGCAGTCTCCAGCACAGCAGCAGCAGATC          1575
                                            |||
 245 ...................................PheSerValGlyIle      249

1576 TCTCAGCAACCTTACCCCCACCAGCGGCCGCCATCCATTTCACAGCAGCA        1625
 244 ..................................................       244

1626 CCAGCAGCAACAGCAGCCACCACCATCGGCCTCAGCAGCATCAGCTTTTTG       1675
     :            |||:::::::
 244 ..........................AspThrIleGlu.....             272

1676 GACACGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTAATTGGGATGT       1725
                                              :::|||
 244 ..........................PheSerValIle                 249

1726 GTGTTTGCAATTGCGGATTATCCAGGAGCAGATGTCTGAAGCAAACTGCT       1775
     |||||||     |||:::::::|||        :::::|||
 250 GlyAlaLysIleAla.....GluSerIleAsnAspCysAspIlePh         263

1776 GGGCACCTGGAAAAGGATAATCCAGGCACACATGCGCGCACTGTTGACCCCA      1825
     :            |||:::::::
 263 eileGlyLeuLysArg.......AspThrIleGlu....               272

1826 CCTTCACGAGTGCGATGCACGCGACCTCTCTGTGAGAGTCAAGTCAGCAGC      1875
 272 ..................................................     272

1876 GCCGTATGCACAGGCAATAACGAAGAAGAAAGAGAGAGTGTTACTCCACACTG    1925
     |||                     :::|||
 273 ...PheAsnLeuAlaSerAsnLysAsnThrThrIleGlyThrIleSerThr      288

1926 G...TTAAACACAGTCTTAAAAAAAAAAAAAAATGGTACCGCGCCCACGGAG      1972
     :  ||||||              |||
 288 pLeuLeuAsnLeuPheValLeuGlySerTrpLysSerProLeuLeuAsnA     305

1973 CCCTTCACTTCCCCAGTGGCCTTCTCCACCAGGAGGAAAGCCATGTTCACAG     2022
     |||||||||||:::|||      ::|||
 305 laLeuHisTyrPro.....PheProSerValGlyPhe.....LeuLysAsp      318

2023 CATATATTATTTCTGTGACTGGATTTGTTGATAGTGACGACAGATGACCTAAA    2072
     |||||||||:||||||::::::|||:::||:   |||   |||
 319 GlnMetValAlaValThrAsnTyrThrAspAlaAlaArgIleTyrLeuGlI      335

2073 ATTAATGG|TTATTTGGCAGTGCCAAAATATACGGGTTATCTATGCCGCA       2122
     ::  |||||||                         |||
```

```
 335 uLysLeuLeuLeuAlaCysGlyAlaThrTyrThrLysAspLeuLysProT 352
2123 GCAACAACGTCTTCATCGTAAACAACAACAACTCGTTTAAAGTAATGAAAAA 2172
     ::|||||||:|:||||||||:::|||||::::|||||||| |||||
2123 hrAsnThrLeuLeuIleAlaAlaSerSerTyrGlyLysTyrGlyValAla 368
2173 GCCAAAGAGTGGAGGAATACCCTGTCAACGCCAGTGGCTTGGCGACAT 2222
     ||||||| |||||||| ||||||||| |||||||:: |||||||
2173 AlaLysValTrpAsnIleProThrValHisHisSerTrpLeu........ 382
2223 TCTTCTCGGCAAACTTCAGCGACTCAGCGACGATTCAGTACGATCGGACTACA 2272
     :::::|||||||||:::|||||||||| :: |||||||
2223 .......TyrSerSerPheLysAsnLeuSerSerGlnAlaThr 395
2273 CGGCAATTC........................... 2280
     |||
2273 hrAspPheProValProLeuAspAspSerTyrMetAspPheIlePhePro 411
2281 .................................AGTCTCAGGATCCA. 2295
     |||||||||
2281 CysProLeuAsnValGlyGlySerPheGluAspThrLeuLysSerSe 428
2296 TTTGGCCCTACC.....................CAGCATTTAGTTTTA. 2322
     |||
2296 erValSerSerIleGlyGlyAsnLysThrAsnGluGlyLeuGlyLysGlu 461
2323 .........................AATCTTTTA 2331
     :::|||
2323 sPArgAsnAspGluGlyTyrSerThrIleThrGlyValAsnArgGluLeuGln 511
2332 GATGTTGGAGAGTT.......... 2346
     |||
2332 AspGluGlyTyrArgLeuGlyIleAspAlaLysLysSerThrAsnThrPr 528
2347 ......CCCTTAAAACGTGTCTGCAGAGTTGTTTGCATGAGTATAAGA. 2385
     ||||| |||::: ||||||::::
2347 oProSerProLeuLeuValGlyThrProSerLysGluSerLeuLysGluA 545
2386 ..................CTACCTCCCAAAACTGAAACAGAAT 2409
     |||||||
2386 laSerSerAspAspGluLeuProValLeuAlaThrLysLeuValAspAsn 561
2410 GAAGTAGCTAAGTCCAGCCTTCTTCCCAAAGAGCCAGAATT.......... 2451
     ::: |||
2410 ValIleLysGluLysSerProLeuSerLeuThrProLysValValPr 578
2452 ...............GAAGAGCTACCACCTCCCACTAAAAAGCTACTCCAGAATTG. 2493
     ::::::||
2452 oSerHisLysGluThrTyrThrAspGluLysLysLeuIleAspGluLeuA 595
2494 .......ACCCCTTTGTCGTTTTCACTGGATTCGAGCCGTCCA 2531
     ||||||
2494 spArgValAsnProLeuAsnSerGlnLeuLeuCysArgSerLysArgLys 611
2532 GGTTCAACAGTATATTAAGAAGCCTCTACAATTCTTGTGCGAGAGGTTCGG 2581
     ::: ::|||||||||| :::
2532 LysAlaSerLysArgValThrTyrIleThrPheThrGlyTyrAspGlyLysSer 678
2582 AGTCTGGACAGAGAGTGGCACACCCTCATTGCCAGCAAGTGAACTCGCAC 2631
     :::::|||||||||||||||| ||||
2582 erAsnProSerLysCysThrHisLeuIleAlaProArgIleLeuArgThr 708
2632 GTGAAGTTCCTGACGGCCGATTTCGTGCGTCGAAGCACATAGTGACGCAGA 2681
     ||||||| :||||:::
2632 SerLysPheLeuCysSerIleProTyrGlyProCysValValThrMetAs 725
2682 GTGGGCTCGGAAGAATGCTTCAGGTGTCAGGAGTTCATTGAATGAGCAGAACT 2731
     :|||||||||||||:|||::: |||||:::
2682 pIleProLeuAsnSerCysLeuLysThrHisIleGluValAspGluLeuPro 742
2732 ACAATTCTCCCGAGATCGCTGAGGCGAGAAGTACTTTCTCTTTCAGCGTTGGAA 2781
     |||||||| ||| :::|||
2732 rValValTyrLeuThrSerLysThrValAlaProGluAsnValProAlaV 792
2782 GAATCCTTAAAACGGGCACACGTTTCT.....CCACTCTTTAAGGCAA 2825
     :: ||||||
2782 alIleLeuSerIleValLysSerAsnGlyValCysSerThrLeuAsnVal 808
2826 AATATTTTTACATCACA.....CCTGGAATCGTGCCCA...AGTCTTTCCACTA 2869
     |||||||| |||
2826 rTyrAsnLysArgLeuAlaArgHisLeuGluGluAspGlyAsn........ 822
2870 TGAAGGCAATCGTAGAGTGTGCAGGAGGAAAGGTGTTATCCAAGCAGCCA 2919
     :::::|||||||||||::: :::
2870 sPAsnAlaSerGlnAsnLysThrIleProLeuGlnAsnTyrAspTrpLeu 855
2920 TCTTTCCGGAAGCTCATGGACGCCACAGCAAGAACTCGAGTTTGTCGGAAAT 2969
     ||||||||||||||::::
2920 IleLysThrValLeuArgGlnGluIleAsp 865
3019 ..................GCCAGAGGCATAGATCGTTCACAATGACGAGAGTTCGTT 3054
     ||| :::::|||
3055 CTGACTGGAGTGGCTCACTCACAAACGTGGAC 3084
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN-MCYj
RX   MEDLINE-91042489; PubMed-2233708;
RA   Laurent B.C., Treitel M.A., Carlson M.;
RT   "The SNF5 protein of Saccharomyces cerevisiae is a glutamine- and
RT   proline-rich transcriptional activator that affects expression of a
RT   broad spectrum of genes.";
RL   Mol. Cell. Biol. 10:5616-5625(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-S288C;
RX   MEDLINE-94378722; PubMed-8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:547-S62(1994)
-!- FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
    COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
    OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
    BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
-!- SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
    COMPLEX.
-!- SUBCELLULAR LOCATION: Nuclear.
-!- SIMILARITY: BELONGS TO THE SNF5 FAMILY.
---------------------------------------------------------------
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation in
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
---------------------------------------------------------------
DR   EMBL; M36482; AAA35062.1; -.
DR   EMBL; X76053; CAA53652.1; -.
DR   EMBL; Z36158; CAA85254.1; -.
DR   PIR; S44551; RGBYS5.
DR   PIR; S39145; S39145.
     SGD; S0000493; SNF5.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN       31    270   GLN-RICH.
FT   DOMAIN       72    132   PRO-RICH.
FT   DOMAIN      272    324   PRO-RICH.
FT   DOMAIN      489    588   ASP/GLU-RICH (ACIDIC).
FT   DOMAIN      714    882   PRO-RICH.
FT   DOMAIN      755    798   ARG/LYS RICH (BASIC).
FT   CONFLICT    564    564   E -> D (IN REF. 1).
SQ   SEQUENCE   905 AA;  102557 MW;  A2B784A648DD1A35 CRC64;
```

```
alignment_scores:
            Quality:   223.50
            Ratio:     1.380
Percent Similarity:    53.465

alignment_block:
US-09-664-641-10 x SNF5_YEAST

Align seq 1/1   to: SNF5_YEAST   from: 1   to: 905

                      Length:   303
                      Gaps:       9
Percent Identity:  27.723

817 CCTGTCTGATTATTTATTGAAGAGGAAGGAAGCAAGAGGAAGAGGAGGA 866
    :||| ::                      ::::::::::::::::::::
 42 ProGlnLeuGlnMetIleGlnHisGlnHisGlnInLeuLeuArgSe 58

867 GGAAGTAGAAAATAGAGAACAAGANTTCTCAGAATGAGGGTAGTACAGATG 916
    ::::::::::::::::||::::|:|||::::||||||:|||||||
 58 rArgLeuGlnGlnGlnGlnGlnGlnGln................ 69

917 AGAAGTCAAAGCCCTGCCAGCTCTCAAGAAGGGTCTCTTCAGGTGACCAG 966
    :: |||      |||            |||||||||   .G 69

 69 InThrSerProGlnInThrHisGlnSerProProProProProGln 85

967 CAGTTTCACCTAAATCAACACTGAAAAATCTAAAGGGGAATTAATGTT 1016
```

```
 86 GlnSerGlnProIleAlaAsnGlnSerAla.................... 95
    :::|||                              :::
1017 TGAIGATTCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
     ::|||                              :::||||||:::
 96 ......ThrSerThrProProProProProAlaProHisAsnLeuHis. 109
1067 GGACCCCGCCCGAAGTCCCACAGTTAGCTGCAGCAAAAGCGACGTCGGCCT 1116
     |||||:::
110 .......ProGlnIle..................... 112
1117 CAGGGAAAGGAGCGCCTGGGTTGATTAACTIGTGTGCCAANTCTCCACCCGT 1166
     ::|||: ||||:::          :::
113 ...GlyGlnValPro.....................LeuAlaProAl 120
1167 CCCAGGTAACACATTTGCCCCCCTGAGGTCCGCGGGGTAATTTAATGGCTG 1216
     |||    ||| ||||||||||:||||:::|||:::
120 aProIleAsn.....LeuProProGlnIleAlaGlnLeuProLeuAlaPhrG 136
1217 GACAAAAACCTCCAAAGTTCNTGAAACGATCAGAAATGATAAGTACCTGGAGT 1266
     |||:::|||       :::||||||    :::::|||:::
136 InGlnGlnValLeuAsnLysLeuArgGlnGlnAlaIleAlaLysAsnAsn 152
1267 CCACGCTGTACGGACACTGAGGAATATATTACTAATAANGCTGACANTTCAGCA 1316
     :::|||:::||||||||:::|||:::||||||:::
153 ProGlnValAsnAlaIleThrValAlaGln.............GlnGl 165
1317 GATGAACCGGCCATCANAAATGTAGCACATATCTTACAGACTCTTTCAGCAC 1366
     |||:::|||||||      :::||||||:::||||||:::
165 nValGlnArgGlnIleGluGlnGlnLysGlnGlnThrAlaGlnThrG 182
1367 CTACGAAAAAATTTAGAACACAGCAGTGAATCAGCAGCCAGCGGACAYACA 1416
     :::|||:::      :::|||:::|||:::
182 LnLeuGluGlnGlnLeuArgGlnLeuLeuValGlnGlnGlnGlnGlnGln 198
1417 AATGCCAATGCACGTGCTGTGTTTAGCCAACGTGAAAGTGACTCCAGAGACA 1466
     |||:::||||||||||:::|||:::
199 LeuArgGlnAsnGlnIleGlnAsnGlyGlnIleGlnPheArgHisHisSva 215
1467 CATGCTCAGCAGCACAGCAGCACAGCAGCAGCACCACCACCACCACCCGG 1516
     |||||||||||||||||||||||||:::
215 lGlnIleGlnGlnGlyGlnIleGlnLeuGlnProGlnInGlnGln.... 230
1517 TTTTTACACCTTCAGCCCCAGCAGATAANTGCAGCTCCAGCAGCAGCAGCAG 1566
     :::|||||||||||||||||||||||:::
231 ....HisGlnGlnGlnLeuGlnGlnProThrAlaGlnThr 245
1567 CAGCAGATCTCTCAGCAAACCTTACCCCCAGCAGCGCCGCCGCCGATCCATTTTC 1616
     |||||||||||||||||||||||||||||:::
246 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 262
1617 ACAGCAGCAGCAGCAGCAACAGCAGCAGCAACAGCAGCACCACCATCG 1657
     ||||||||||||||||||||||||CCTC 1657
262 nGlnGlnGlnGlyGlnIleGlnLeuProGlnSerGlnInGlnValProG 279
1658 AGCACACATCAGCTCTTTTGGACATGATCCAGCAGTGGAGATTCCAGGAAGA 1707
     ::|||||||||||||:::
279 InValArgSerMetSerGlyGlnProThrAsnValGlnGlnThrIle 295
1708 GGCTCTTA 1716
     |||
296 GlyGlnLeu 298
```

```
seq_name: SwissProt_40:YKF4_YEAST

seq_documentation_block:
ID   YKF4_YEAST     STANDARD;     PRT;    738 AA.
AC   P35732;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Hypothetical 84.0 kDa protein in NUP120-CSE4 intergenic region.
GN   YKL054C OR YKL308.
```

OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomyces;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378723; PubMed=8091862;
RA   Rasmussen S.W.;
RT   "Sequence of a 28.6 kb region of yeast chromosome XI includes the
RT   FBA1 and TOA2 genes, an open reading frame (ORF) similar to a
RT   translationally controlled tumour protein, one ORF containing motifs
RT   also found in plant storage proteins and 13 ORFs with weak or no
RT   homology to known proteins.";
RL   Yeast 10:563-568(1994).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X75781; CAA53418.1; -.
DR   EMBL; Z28054; CAA81890.1; -.
DR   PIR; S37876; S37876.
DR   PIR; S39184; S39184.
DR   SGD; S0001537; YKL054C.
KW   Hypothetical protein.
FT   DOMAIN   382   718   GLN-RICH.
FT   DOMAIN   574   580   POLY-ALA.
FT   DOMAIN   708   716   POLY-ALA.
SQ   SEQUENCE   738 AA; 83973 MW; 05734E2D0B7389AC CRC64;

alignment_scores:
        Quality: 213.00
        Ratio: 1.275
Percent Similarity: 51.543   Percent Identity: 26.543

alignment_block:
US-09-664-641-10 x YKF4_YEAST

Align seq 1/1   to: YKF4_YEAST   from: 1   to: 738
                                            Length: 324
                                            Gaps: 11

835 GAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGTAGAAAATGAGGA 884
     :::|||
199 GluLysGluGlnGluGlnGlnGlnGlnGluGlyHisAsnAsnLysSl 215

885 ACAAGATTCTCAGAATAGAG........................ 903
     :::||| :::|||
215 uGluHisLysGlnIleGluGlnProSerLeuSerSerLysLysThrhrS 232

904 ........GGTAGTACACGATGAGAAGTCAAGCCCTGCCCAGCTCTCAA 942
     ::||| ::||| |||
232 erArgThrSerAlaSerGlnProLysLysMetSerTrpAlaAlaIleAla 248

943 GAAGGGTCTCCTTCAGGTGACCAGCAGTTT...TCACCTAAATCCAACAC 989
     :::|||||||||| :::|| |||:::||:
249 ThrProLysProLysAlaValLysLysThrGluSerProLeuGluAsnVa 265

990 TGAAAAAATCTAAAGGGGAATAATGTTCATGATTCTTCAGATTCCATCAC 1039
     :::||| ||| |||
265 1AlaGluLeuLysLysGluIleSerAspIleLysLysAspSerGlnLys 282

1040 CCGAAAAAACAGCAGAGAAATTTAAACTGGACCCCGGCCGAAGTCCCACAG 1089
     ||| :::||| :::|||
282 erGluAlaSerGlnGluLysLysValAsn............GluGln 292

1090 TTAGCTGCAGCAAAACGCAGGCGGTGCCTCAGGAAAGAGAGCCTGGGTTGAT 1139
     ::::|||||||| |||:: |||:::||:
293 GluThrSerAlaGlnGluGlnGlnGlnThrAlaGluProSerGluG 309

1140 TAACTTGTCGTGCCAATGTCCCACCGTCCCAGCGTAACAATTTTGCCCCTG 1189
     |||:::
309 uAsnGlu...AspArgValProGluValAspGlyGlu.........G 321

1190 AGGTCCGGGGTAATTTAATGCTCTGACAAAAACCTCAAAGTTGTCGAA 1239
     :::|||||| :::||
321 luVal...Gln.....................GluGluAlaGlu 327

1240 AGATCAGAAATGATAGCTACCTGGAGTCCAGCTGTACCGGACACTGAGGAA 1289
     :::||| :::|||::: |||:::
328 LysLysGluGlnValLysGlnGlnThrAlaGluGluGluGluGluGl 344

1290 TATTACTAATAAT.......GCTGACATTCACGACATGAACCGGCCATCAA 1333
     : ::||| ||| :::|||:::
344 nGluGlnAspAsnValAlaAlaAlaProGluGluGlnValThrValGluG 361

1334 ATGTTAGCACATATCTTACAGACTCTTTCAGCACCTTACCGAAAAAATTAGAA 1383
     :::||| :::||::: |||
361 luLysValGluIleIleSerArgLeuAlaValIleSerGluProGlnAla 377

1384 CAGCAGGGTGAATCACAGCCAGGAGACAGCAGGAGACAAAATGCCAATGCACAGTGCT 1433
     ::||| |||:::
378 AsnThrValProGlnProGlnGln. 385

RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous-nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L13200; AAA28193.1; -.
DR   PIR; S44898; S44898.
DR   WormPep; 2K1236.3; CE00532.
KW   Hypothetical protein.
SQ   SEQUENCE   1000 AA; 113419 MW; D230C9A6F8492EB2 CRC64;

alignment_scores:
              Quality: 204.50
              Ratio: 0.492
Percent Similarity: 41.352

alignment_block:
US-09-664-641-10 x YO83_CAEEL

Align seg 1/1 to: YO83_CAEEL   from: 1   to: 1000

                                    Length: 1006
                                    Caps: 47
Percent Identity: 19.682

139 AGGGGTACGGATCTCTGGCCCCCTGCCCTTCAAAATCACTCTAGGCACAG 188
    ||| ||| ||| ||||| | ||||||||||||
104 ArgMetLysAspThrSerLeuProProLeuGlnAsnProSerAsnGlyGl 120

189 GTTATTCACCTTGTCCAAGCCTGAAAAACGGAAGGAAGTTTCCTACAATGC 238
    ||| ||| ||||||| :|| ||||||||||||
120 yThrTyrLeuPheAsnLeuIle............ThrValPheProSerIleA 134

239 ACTAGGCCTCACACATAATCTCAGAGGATGGGACAANTCCAGAGGTGGGAG 288
    ||| :|| ||| ||||| ||| |||||||||||||
134 laLeuSerPheAsnThrSerThrArgGlyAspThrGlyProSerIleSer 150

289 AAGCTCGGGAAGTCTTTTGACTTACCTGTGTAAAGCCCTTCTGGGTGATT 338
    :||||||||||::| |||||||| ||||||||||||||:::
151 IlePheSerPheSerLeuSerValLeuPhePheSerLeuPheProGlnHi 167

339 CTGCCCGTTCAGTCTGGGACTCTTCTCGCCAGTAAATGGTTTTTCTCCAGA 388
    ||| |||::: ||||| ||||||||||||||:::
167 sLysAsnIleCysAlaAargAlaArgpCys................ 176

389 ATCATGTCAGATTTTTTTTGGAATCACTGCCGCCTTTCTCAGGGTGTTG 438
    ||| ||| ||::||||||| :|| || :|| :|:
177 ....ArgProThrArgSerLeuSerPheSerLeuArgPheIle 190

439 ATACAAGCTGGAGCTCTTTGTTGGAGTGTCTTCCAGAGCTCTCCAGGAGA 488
    ::::: |||| ||| |||::| ||||||||||||
191 MetSer..........................SerGluProAlaSerSerThrGl 201

489 G...GTAGGGAAGGGAGCTTGTCCAGCAGAAGTTGGAAGCACAGAGATC 535
    ::: |||||||||| | |||||||||||||| |||:::
201 uLysValProGluGluProHisProHisSerIle...LysHisLysPheG 217

536 ATCTGGCCTTCTTCTGACCCCGGTATTGATGCAGGCTGAGGCCTCTGTTGTA 585
    |||: :|||| ||| ||:|| :||:::|| :|:
217 lnGlyProGlnPhe........ValIleProArgAlaLeuSerAspTyrVal 231

586 ATGTGCTGGGTGTCATCTGAAGACAGAAGTGCCCTGTGGGTTGGTTAC 635
    ||||: ||| |||||||||||||| ||||| ::|
232 LeuAsnValAsnAsnGlnThrProGluSerTyrGluLysAlaLeuAsnAl 248

636 GTTCTTATGCGGGGGAGAGTTGCCAGCTAACCCTCAATAAGAAATTGCACGCATT 685
    ::: ||||||||||| ||||||||||||||||:::|
248 aLysThrTyGlYArgAspAspTyrIleThrLeu........CysLeuHisV 262

686 TGATTGTT............CCAGAGCCAAAGGGGGGGAAATAC 717
    :: ||| :::||||||||||||||||
262 alThrSerLeuCysThrGluTyrGlyProLeuAspAspGlyProGluTyr 278

718 GAA.......TGTGCTTTAAAGCGAGGCAAGCTATTAAAATTGTG....... 753
    ||| |||||||||||||||||||||||||||:::
279 ValLeuLeuCysAspSerArgAlaArgValAspLysLeuIleGluAspIle 295

753 753
295 uValLysLeuLeuGluIleAspThrAspTyrValGlnLeuGluLeuHisG 312

754 ...........ACTCCTGACTGGGGTTCTGGATTGGGTA 780
    ::: ||||||| ||||||||||||
312 lyGlyLysArgLeuHisLeuGlnLysProAspProAlaValLeuArgAspIle 328

781 TCAGAGAAAACCAAA........AAGGACGAAGCATTTTATCATCCTCG 821
    ||||||||||| :::|||||||||||||
329 AlaTyrLysGlnSerAsnGlyGlySerAspLysSerAspPhePheLeuGluMetL 345

822 TCTGATT.............ATTTATCAACAGAGGAAGAAG 847
    ||||||| ||||:::|||||||||||||
345 sLeuValProSerGluSerMetLysAlaLysIleMetLysGlnGlnGluG 362

848 AGGAAGAG.................GAAGGAGGAGGAAGTA 873
    ||||||| :::|||||||||||||
362 luGluGluLysAlaLysArgLysHisGlyGlnThrArgGlnThrTyr 378

874 GAAAAATGAGGAACAA............GATTCTCAGAATGAGGGTAGTAACAGA 914
    ::::::||||| |||||||||||:||||||
379 HisGlyGlnHisGlnAlaMetAsnAspGlnSerSerSerValPr 395

915 TCAGAAGTCAAGCCCCTGCCAGCTCTCAAGAGGGTCT........ 951
    ::: |||||||||||||||||||||||||
395 oSerThrSerSerProSerCysSerSerGluAlaAsnArgLysGluMetG 412

952 ...........CCTTCAGTGACCAGCAGTTT.......... 972
    ||||||| :::|||||||||||||
412 luThrValArgGluProAlaGlyProSerGluLeuMetArgAlaIleAsn 428

973 ....TCACCTAAATC 983
    ::: ||||||||
429 AlaProValAlaProAlaValValIleLysIleGluGluThrProAlaVal 445

984 CAACACTGAAAAAATCTAAAGGGGAATTAAGTTTGATGAATTCTTCAGATT 1033
    : |||||||||||||||||::||||||
445 aLeuProGluGluProAlaAspGlu...ThrLeuMetAspAspSpGluMetProS 461

1034 CA.................................. 1035
     ||
461 erLeuThrValGluAlaProSerGluValAlaSerPheGluLeuGln 477

1036 .......TCACCCGAAAAAACGAGAGAGAAATTTAAAACTGG..... 1076
           |||||||:::: |||||||||||||
478 ProSerProGlnValProGlnAlaSerIleGluGlyProSerGlnGlnIle 494

1077 CGAAGTCCCACAGTTAGGTCGCAGCAAAAACGACAGGGTGCCTCAGGGAAGG 1126
     ::: |||||||| |||||||||:::
494 nGlnIleProGlyThrSerGlnIleLysArgGlnValAlaArgGlySerA 511

1127 AGCCTGGGGTTGATTAACTTGTGTCCCAATGTCCCACCCGTCCCAGGTAAC 1176
     :::::||||| ||||:::::|||||||
511 rgThrAsnMetIleSerTyr...HisAspLeuProProGlyThrGlyAsn 526

```
1177 ATTTTGCCCCTCGAGGTCCGGGGTTAATTTAATGGCTGCTGGACAAAACCT 1226
       ||||||                          ||||||||  ||||
 527 ...AlaProPro....................MetAlaCysProGlnValTh 536

1227 CCAAAGTTCTGAAAGATCAGAAATCATACACCTACCTGGAGTCCAGCTGTAC 1276
       :::                              :::
 536 rLeuLysLeuGluLysAsn........................ValProPheGluAlaLysIleA 550

1277 GGACACTGAGGAATATATACTAATAAT......CCTGACATTCAGCAGATG 1320
       ||:::|||                         |||:::|||:::|||
 550 rgAlaValAlaGlyTyrThrArgLysProIleSerGluValGlnLysMet 566

1321 AACCGGCCCATCAAATGTAGACACATATCTTACAGACTCTTCAGCACCTAC 1370
       |||||||||:::                      |||:::::::
 567 ..ArgProSerAspLeuGluSerIlePheHisSerIle...... 578

1371 CAAAAATTTAGACACAGCAGGTGAATCACACCCAGGGACATACAAATC 1420
       :::::                            ||:::|||::::|
 578 ............................................. 578

1421 CCAATCACTTGCTCTTTAGCCAAAGTGACTCCAGAGACACACATG 1470
       :::                             :::
 579 ........CysIleAlaSerValGlnArgIleLysArgArgAsnGluLeu 592

1471 CTACAGCAGCAGAGGCCCCAGCAGCAGCACCACCGGTTTT 1520
       ::|||||||||||                    |||:::
 593 ValGlnGlnLeuGlnLeuIleAsnAlaClrs...CysLysSerProThrMe 609

1521 ACACCTTCAGCCC. 1545
       ::::||||||||                     CAGCAGATAATG 1545
 609 ................................ 609

1546 .CACTCCAGCAGCAGCAGCAGCAGCAGATCCTCAGCAACTTTACCCC 1593
       |||||||||||||||||                |||:::
 626 snCluLtlCGlutysIleaspArgGluGlnIleLeuProGlnTyrMet 642

1594 CAGCACCCGCGCCATCCATTTTCACGAGCAGCAGCAGCAACAGCAGCC 1643
       :::::||||||||                     |||:::
 643 AsnMetProProMetPro....................ProGlnGlyGlnGlnArgLeuPr 656

1644 ACCACCATCC. 1672
       ||||||||                          CCTCAGCAGCATCAGCTTT 1672
 656 oProProAlaTyrProProGlyIleLeuProProGlnGlnAsnArgGlnG 673

1673 TTTCGACATGAATCCAGCAGTGAGATTCCAGAAGCAAGCAGGTTCTTATTGGGA 1722
       :::                              :::
 673 InGlyValProProGlnPheGlnArgSerProGln...PheMetIleGly 688

1723 TGTGTGTTTGCAATTGCGGAATTATCCAGAGCAGAGTCTGAATAAGCAACT 1772
       |||                              |||
 689 ..ProAspGlyGlnAr 693

1773 GCTTCGCCACCTGGAAAAAGGATAATTCAGCCACCATGCGGCGGCGACTGTTGACC 1822
       |||||||                          :::
 693 g.....................TyrAlaHisProTyrMetGlnLeuP 702

1823 CCACCTTCACGAGTCGAGATTCCTCGTGAGAGTCAAGGTCAGC 1872
       ||::                             |||:::||
 702 roAsnSerAsnGlnArg.....AlaArgIleLeuAsnThrSerSerValGln 717

1873 AGCGGCGTATGCACAGCGCAATAAGAAAGAAGAAGAGATGTGTACTGCACA 1922
       :::                              :::
 718 Pro.......SerGluGlyValArgGlyAsnArgLeuValLysIleGluAlaMe 732

1923 CTGCTTAAACACAGTCTTA............AAAAAAAAATGGTAC 1960
       ::::|||                          :::
 732 tAlaMetAsnMetLAlaGlnLeuAsnProProArgProProGlnP 749

1961 CCCCACCACGACGCCCTTCACTCCCAGTGCGGCCTTC.....CCACCAGGA 2004
       |||||||||||||||||                |||
 749 roProHisArgAlaLeuGlnGlyGluLeuGlnPheLeuArgProGlyAla 765

2005 GGAAAGCCATGTTCACAGCATATATTATTCTGTGACTGGATTGTTGATAG 2054
```

```
 766 ProAspProCysAsn.................... 770
       |||||||:::
2055 TGACAGAGATGACCTAAAATTAATGCCTTATTGGCAGGTGCCAAATATA 2104
                                        :::
 770 ............................PheArgProAspSerLysGluT 778

2105 CGGGTTATCTATGCGCACCACACACAGTCCTCACCTGTAAAGAACAACT 2154
       :::||||||                         ::::||||||
 778 hr..........TyrAsnAsnThrTyrValThrValAlaSerProAla 790

2155 GGTTTA....................... 2174
                                        AAGTATGAAAAAGC
 790 ............................... 790

2175 C..... 2188
       |||:::::||                         AAAGAGTGGAGGA
 791 ThrLeuThrAsnSerIleIleIleProTrpHisPheProProTyrGluLysSe 807

2189 TACCCTGTGTCAAGGCCCAG.............. 2208
       :::||||||
 807 rGlyArgLeuAsnValSerAsnThrIleLysAlaIleLeuAsnGluTyrArgL 824

2209 TGGCTTGGcGACATTCTTCTGGGAAACTTTGAGGCACTGAGGCAGTTCA 2258
       :::||||||                         |||:::
 824 euLeuCysAsnSerArgGlnAlaAspProAlaSerPheLeuGluPheTyr 840

2259 GTATAGTCGGTCTACACGGGCATTCAGCTGTCTGCAGGATCCA........ 2295
       :::||:::|||                       |||
 841 PheLeuGlyAspProMetProHisPheAsnLysIleLeuSerIleAlaAs 857

2296 TTTGCCCCTACCCACCATTTAGTTTAAAATCTTTTAGAT.... 2334
       :::||||||                         |||
 857 pTyrAsnMetTyrLeuSerArgArgArgCysAspGluAlaAspValLysI 874

2335 .GCTTCGAGAGTTCCCTTA....................... 2352
       |||                              :::||||||
 874 leHisArgMetSerHisSerAspGlnLeuGlnLeuGlnTyrLeuLeuGluGluLeu 890

2353 .................................. 2364
                                        AAAGGTCTTGGA
 890 ................................ 890

2365 GAGTTGTTGATGAGT.................... 2396
       :::||:::
 891 GlnSerAspGluSerAsnValGlnGluLysTrpLysThrPheTyrArgIleMe 907

2397 ACTTCAAAACAGAATGAA......... 2412
       :::||||||
 907 tGlnIleAspLeuProLeuAsnAsnGluProProArgIleLeuLeuLeuPro 924

2413 ......GTAGCTAAATGTCCAGCCTTCTTCCCAAAAGAGCCAGAATTGAA 2454
       :::||||||:::                      ::||||||:::|||
 924 erSerLeuAspIleGlyArgProValValAspArgGlyLysLysSerIle 940

2455 GACGTACCACCTCCCACT 2472
       ||||||||                          ATAAGACTACCTCCCAA
 941 AspGlnValMetAsnHisIleHisAsnMetHisSerGlnArgProProSe 957

2457 .................................. 2412
                                        :::
 957 rMetGlyAsnSerSerThrSerSerGluAlaSerSerThrSerProThrA 974

 974 snAlaAlaThrAlaThrSerSerProAlaSerAsnArgProThrThrSer 990

 990 ............................... 990

 991 ThrAlaGlnProThr 996
```

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-i., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
-!- TISSUE SPECIFICITY: UBIQUITOUS.

CC   -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D83783; BAA12112.1; -.
DR   MIM; 300188; -.
KW   Hypothetical protein.
FT   NON_TER        1
FT   DOMAIN       599    602       POLY-SER.
FT   DOMAIN      1201   1207       POLY-GLY.
FT   DOMAIN      1998   2124       GLN-RICH.
FT   DOMAIN      1998   2023       POLY-GLN.
FT   DOMAIN      2028   2033       POLY-GLN.
FT   DOMAIN      2037   2070       POLY-GLN.
FT   DOMAIN      2090   2097       POLY-GLN.
SQ   SEQUENCE   2124 AA;  237207 MW;  255FB9419BC39F42 CRC64;
```

```
alignment_scores:
        Quality:        200.00
        Ratio:          0.952
Percent Similarity:    42.000

alignment_block:
US-09-664-641-10 x Y192_HUMAN

Align seg 1/1 to: Y192_HUMAN    from: 1  to: 2124
                                   Length:    500
                                   Gaps:       22
Percent Identity:  23.200
```

```
658  CTAACCTCAATAAGAAAATGCACGCATTTGATTGTTCCAGAGCCA.........  702
        ||||  ||||::::::: ||||||:::||||||||||||
586  LeuProLeuProLysGlnThrArgAspValIleThrCysProGlnGlnGl  602

703  ...................................TACGAATGTGCTTG  727
                             AAGGGGGGAAA            ::::
1602 ySerLeuIleAspThrLysGlyAsnLysIleAlaGlyPheAspSerIIeP  1619

728  TAAAGCGAGCAAGTAATTAAAATTGTTGACTCCTGAC............TGG  765
        :|||||:::: |||||||:::::::::||| |                 |||
1619 heLysLysGluGlyLeuGlnValSerThrLysGlnLysIleSerProTrp  1635

766  GTTCTG..........................................  771
        |||
1636 AspLeuPheGluGlyLeuLysProSerAlaProLeuSerTrpGlyTrpPh  1652

772  ......GATTCCGTATCAGAGAAAAACCAAAAAAGGACCAA.........  804
        :|||:::           |||:::: |||||
1652 eGlyThrValArgValAspArgArgValAlaArgGlyIleGlnGlnInA  1669

805  .....GCAATTTTATCAT.............CCTCGTCTCGATTATTTATGAA  837
              ||||||||||                 ||||||
838  GAGGAA.......................................GAAGAGGAGGAGGAGGAGGAGGAGGA  869
        |||                                    |||||:::|||||
1669 rgLeuLeuLeuTyrHisThrHisLeuArgProArgProArgAlaTyrTyr  1685
```

```
1686 LeuGluProLeuProLeuProProGluProGluAspGluGluProAlaProTh  1702
1702 AGTAGAAAAATGAGGAAAAAGAATTCAGAAGATGAGGTAGTACAGATGAGA  919
        :::           |||||||||||:::      ::::||||||||||:::
870  AGTAGAAAAATGAGGAAAAAGATTCAGAAGATGAGGTAGTACAGATGAGA  919

1702 rLeuLeuGluProGluLysLysAlaProGluProProLysThrAspLysSP  1719
        :::||||||||:::|||    |||:::
920  AG...TCAAGCCCTGCCAGCTCTCAAGAA.............GGGTCTCCTTCA  957

1719 roGlyAlaAlaProProSerThrGluGluArgLysLysLysSerThrLys  1735
        :::|||         |||:::||||||:::||||||        |||
958  GGTGACCAGCAGTTTCACCTAAAATCCAACACTGAAAAA...........  996

1736 GlyLysLysArgSerArgSerGlnProAlaThrLysThrLysAspTyrGlyMetGl  1752
        |||:::|||  ||||| |||:::|||||||||||||
997  ...............TCTAAAGGGGAATTAATGTTTGATCATTCTTCAG  1030

1752 yProGlyArgSerGlyProTyrGlyValThrValProProAspLeuLeuH  1769
        |||           |||:::||||||||||||||:::
1031 ATTCATCACCGGAAAAACAGGAGAGAAATTTAAACTGGACCCCGGCCAA  1080

1769 isHisProAsnProGlySerIleThrHisLeuAsnTyrArgGlnGlySer  1785
        :::|||||||:::||||||:::::::
1081 GTC...............CCACAGTTAGCTTGCAGCAAAACGC......  1107

1786 IleGlyLeuTyrThrGlnAsnGlnProLeuProAlaGlyGlyProArgVa  1802
        |||:::::::           |||::::::||||||||||
1108 ...............AGGCTGCCTCAGGGAAAAGGACCCTGGGTTGA  1138

1802 lAspProTyrArgProValArgLeuProMetGlnIleLysLeuPro....  1816
        ||||||||||:::||||||:::::::
1139 TTAACTGTGTGCCAATGTCCCACCCGTGCCAGGTAACACATTTGCCCCT  1188

1817 ...................ThrArgProThrTyrProGly...ValLeuProThr  1827
        ::::||||||         |||:::::::
1189 GAGGTCCCGGGTAAATTTA.............................  1206

1828 ThrMetThrGlyValMetGlyLeuGlyGlnSerSerValHisGlnMet  1844
        :::           |||
1861 InLeuGlnSerGlnGlyMetLeuGlyGlnSerSerValHisGlnMet  1877

1231 ..................................AGTTCTGAAAGATCAGAAATG  1251
                                          ::::||:
1861 rValSerHisValGlyLeuGlnHisThrGlyProAlaGlyThrMetV  1911

1844 lTyrArgGlnGlnGlnProAlaValProGlnGlyGlnArgLeuArgGlnC  1861
        ::::::|||||||||           ::::::
1207 ...........AGGCCTGCTGGACAAAACCTCCAA..........  1230

1861 InLeuGlnSerGlnGlyMetLeuGlyGlnSerSerValHisGlnMet  1877

1297 ......AATAArgCTGACAATTCAGCAG.............ATGA  1321
           ||||||:::||||||||           |||
894  rValSerHisValGlyLeuGlnHisThrGlyProAlaGlyThrMetV  1911

1378 ThrProSerSerThrGlyLeuGlnThrSerGlnGlyThrThrProTy  1894
        :::||||||           |||
352  ThrProSerSerThrGlyLeuGlnThrSerGlnGlyThrThrProTy  1894

1354 ...ACTCTTTCAGCACCTACCAGAAAAAATTTAGAACAGCAGGTGAAT  1395
        |||||||:::||||||           |||:::|||
1911 alProProSerIIrSerGlnInProTyrGlnSerThrHisProSerThr  1927

1322 ACCGGCCATCAAAATGTAGCACAATAGCTTACAG........  1353
        |||||||
1944 AsnProThrLeuValAspProThrArgHisLeuGlnAlaArgProSerGl  1944

1396 ................CACAGCCAG.  1404
        :::||||||
1961 InArgPheSerHisGlnThrLeuGlnThrProMetIleSerThrMet  1977

1944 yTyrValHisGlnGlnAlaProThrTyrGlyHisGlyLeuThrSerThrG  1961
        |||:::

1404 ................ACTGTTTCAGCACCTACCAGAAAA  1404

1978 ThrProMetSerAlaGlnGlyValGlnAlaGlyValAlaArgSerThrAlaIl  1994
        :::||||||           |||:::
1405 CAGGGACATACAAATGCCAAATGCAGT  1430
```

```
1431 GCTGTTTAGCCAAGTGAAAAGTGACTCAGAGACACAATGCTACAGCAGC 1480
     :||||      :::||||
1994 cLeuProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2011

1481 AGCAGCAGCCCAGCAGCAGCAGCAGCAGCAGCAGCACCCGGTTTTACCTTCAG 1530
     |||||||||  ||||||||||||||||||:::
2011 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnTyrHisLeuArg 2027

1531 CCCCACCAGATAATGCAGCCTC.............CAGCAGCA 1559
     |||||        ::::||||||||                  ||||||||
2028 GlnGlnGlnGlnGlnlleLeuArgGlnGlnGlnGlnGlnGl 2044

1560 GCAGCAGAGCAGAGATCTCCAGCAACCTTACCCCAGCAGCGCCGGCATC 1609
     |||||||||||| |||||||  :::||||||||||||||||:::
2044 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2061

1610 CATTTTCACAGCAGCAGCAGCAGCAGCAGCACCACCATCGCCTCAG 1659
     |||||||||||||||||||||||||||||  |||||::::|||||
2061 lnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnAlaAlaProProGlnProGln 2077
```

seq_name: SwissProt_40:FXP2_MOUSE

seq_documentation_block:
```
ID  FXP2_MOUSE        STANDARD;    PRT;    714 AA.
AC  P58463;
DT  01-MAR-2002 (Rel. 41, Created)
DT  01-MAR-2002 (Rel. 41, last sequence update)
DT  01-MAR-2002 (Rel. 41, last annotation update)
DE  Forkhead box protein P2.
GN  FOXP2.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=C57BL/6; TISSUE=Lung;
RX  MEDLINE=21347947; PubMed=11359962;
RA  Shu W., Yang H., Zhang L., Lu M.M., Morrisey E.E.;
RT  "Characterization of a new subfamily of winged-helix/forkhead (Fox)
RT  genes that are expressed in the lung and act as transcriptional
RT  repressors.";
RL  J. Biol. Chem. 276:27488-27497(2001).
-!- FUNCTION: Transcriptional repressor that play an important role in
-!-    the specification and differentiation of lung epithelium. May play
-!-    important roles in developing neural, gastrointestinal and
-!-    cardiovascular tissues.
-!- SUBCELLULAR LOCATION: Nuclear (Probable).
-!- TISSUE SPECIFICITY: Highest expression in lung. Lower expression
-!-    in spleen, skeletal muscle, brain, kidney and small intestine.
-!- DEVELOPMENTAL STAGE: Expressed in developing lung (only distal
-!-    epithelium), neural, intestinal and cardiovascular tissues.
-!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
-!- SIMILARITY: CONTAINS 1 C2H2-TYPE ZINC FINGER.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```
DR  EMBL; AF339106; AAK69651.1; -.
DR  MGD; MGI:2148705; Foxp2.
DR  PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR  PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR  PROSITE; PS50039; FORK_HEAD_3; FALSE_NEG.
DR  PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR  PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW  Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW  Nuclear protein.
```

```
FT  DNA_BIND    503    593    FORK-HEAD.
FT  ZN_FING     345    370    C2H2-TYPE.
FT  DOMAIN       53     56    POLY-GLN.
FT  DOMAIN      123    126    POLY-GLN.
FT  DOMAIN      131    136    POLY-GLN.
FT  DOMAIN      152    191    POLY-GLN.
FT  DOMAIN      200    208    POLY-GLN.
FT  DOMAIN      222    230    POLY-GLN.
SQ  SEQUENCE    714 AA;    79820 MW;  BCDFB80E28398609 CRC64;
```

alignment_scores:
```
            Quality:   198.50           Length:   352
            Ratio:     1.196            Gaps:      14
Percent Similarity: 47.159    Percent Identity: 25.284
```

alignment_block:
US-09-664-641-10:
US-09-664-641-10 x FXP2_MOUSE

Align seq 1/1   to: FXP2_MOUSE   from: 1 to: 714

```
 850 GAAGAGGAAGAGGAGGAGGAGGAGTACAAAAATGAGGAACAAGATTCTCAGAA 899
     ::::||||:::              ||| |||||::::|||::|     :::||||
   3 GlnGlnGluSerValThrGluThrIleSerAsnSerSerMet....AsnGlnAs 18

 900 TGAGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGTCTTCCAAGAAGGGT 949
     | |||||    ::::||||:::              ||||||:::||::::|||| |||
  18 nGlyMetSerThrLeuSerGlnLeuAspProAlaGlySerArgAspGlyA 35

 950 CTCCTTCAGGTGACCAGCAGTTTTCACCTRAAATCCAACACTGAAAATCT 999
     |||||||||                       ::::::::|||(|||| |||
  35 rgSerSerGlyAsp................ThrSerSerGluValSer 45

1000 AAAGGGGAATTAAATGTTTGATGAATTCTTCAGATTCATCACCGGAAAAACA 1049
     |||||:::              :::              |||||||
  46 ThrValGluLeuLeuHisLeuGlnGlnGlnGlnAla................ 57

1050 GGAGAGAAAATTRAAACTGGACCCCGGCGCGAAGTCCCACAGTTAGCTGCAG 1099
     :::||||||                           ||| |||| ||||
   3 GlnGluSerValThrGluThrIleSerAsnSerSerMet....AsnGlnAs 18

1100 CAAAACGCAGGCTGCCTCAGGAAAGGAGCCTGGGTTGATTAACTTGTGT 1149
     |||||||                 :::      ||||| 
  61 laArgGlnLeuLeuGlnGlnGlnThrSerGlyLeu................ 73

1150 GCCAAATGCTCCCACCCGTCCCAGGTAACATTTTGCCCCCTGAGGTCCGGGG 1199
     :::              ||||
  73 ................. 73

1200 TAATTTAATGGCTCGGGACAAAAACCTCCAAAGTCTCGAAAGATCAGAA 1248
     :::::              :::||||||:::||::::
  74 ...............LysSerProLysSerGlnLysGlnArgP 84

1249 ...........ATGATAGCTACCTGGAGTCCACTGTACGGACA 1281
     :::||||                       |||||| |||
  84 roLeuGlnValProValSerValAlaMetMetThrProGlnValIleThr 100

1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGATGAACCGGCCATC 1331
     :::||||:::              ||||| ||||| |::::
 101 ................ProGln 102

1332 AAAATGTAGCCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATTTAG 1381
     :::||||      :::::||||||::::
 102 nGlnMetGlnGlnIleLeuGlnGlnGlnValLeuSerProGlnGlnLeu 119

1382 AACAGCAGGTGAATCACAGCCGCAGGGACAATACAAAATGCCAATGCAGTG 1431
     :::||::::        |||||:::              ||||
 119 lnAlaLeuLeuGlnLeuGlnGln................AsnGlnAs 128

1432 CTGTTTAGCCAA..........................GTGAAAGTGACTCCAGA 1460
     ::::::||||                              ||| :::
 129 MetLeuGlnGlnGlnIleLeuGlnGlnGluPheTyrLysLysGlnGlnGluGl 145
```

```
1461 GACACAC.......ATGCTACAGCAGCAGCAGCCAGCAGCCAGCAGCCAGCAGCAGCAGC 1504
          :||     |||||||||||||||||||||||||||||||||||||||||||
145 nLeuHisLeuGlnLeuLeuGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 162

1505 AGCAGCACCGGTTTTACACCTTCAGCCCCAGCAGCAGATAATGCAGCTCCAG 1554
          :||:||          ||||||||||||||||||| |||
162 lnGlnGlnGln.........GlnGlnGlnGlnGlnGlnGln 174

1555 CAGCAGCAGCAGCAGCAGAGNTCTTCAGGCAACCTTACCCCCAGCAGCCGCC 1604
          |||||||||||||||||||      :::||||||      |||||
175 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln.GlnGl 191

1605 GCATCCATTTTCACAG.........CAGCAGCAGCAGCAACAGCAGCCAC 1645
          ||||||| ::|||                ||||||||||||||||
191 nHisProGlyLysGlnAlaLysGlnGlnGlnGlnGlnGlnGln 208

1646 CACCATCGCCTCAGCAG..................CAT 1665
          ||||||| ::|||
208 lnLeuAlaLeuGlnHisLeuValPheGlnGlnGlnLeuLeuGlnMetGln 224

1666 CAGCGTTTTTGGACATGATCCAGCAGTGGAGATTCCAGAGAAGGCTTCTT 1715
          |||||||||||||||| :::|||||||| ||||:::|||||:::||:|||
225 GlnLeuGlnGlnLeuGlnHisLeuLeuSerLeuGlnAArgGlnGlyGlyLeuIl 241

1716 ATTG.......GGAGTGCTTGTTTGCAATTGCGGATTATCCAGAGC 1753
          :::        ||||:::||||||::||:::          :::||:::
241 eSerIleProGlyIleAlaAlaLeuProValGlnSerLeuProGlnAla 258

1754 AG...AIGTCTGATAAGCAACTGCTGGCCACCTGGAAAAGGATAATCCAG 1800
          :::    :::::||          |||||||||||||||||
258 laGlyLeuSerProAlaGluIleGlnGlnLeuTTPLysGluValThrGly 274

1801 GCACAT 1806
          :::|||
275 ValHis 276
```

```
seq_name: SwissProt_40:FXP2_HUMAN

seq_documentation_block:

ID   FXP2_HUMAN        STANDARD;        PRT;        715 AA.
AC   O15409;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update).
DE   Forkhead box protein P2 (CAG repeat protein 44) (Trinucleotide repeat-
DE   containing gene 10 protein).
GN   FOXP2 OR CAGH44 OR TNRC10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT SPCHI HIS-553.
RC   TISSUE=Brain;
RX   MEDLINE=21470412; PubMed=11586359;
RA   Lai C.S.L., Fisher S.E., Hurst J.A., Vargha-Khadem F., Monaco A.P.;
RT   "A forkhead-domain gene is mutated in a severe speech and language
RT   disorder.";
RL   Nature 413:519-523(2001).
RN   [2]
RP   SEQUENCE OF 1-304 FROM N.A.
RX   MEDLINE=97369492; PubMed=9225980;
RA   Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA   Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT   "cDNAs with long CAG trinucleotide repeats from human brain.";
RL   Hum. Genet. 100:114-122(1997).
RN   [3]
RP   SEQUENCE OF 1-86 FROM N.A.
RA   Minx P., Hinds K., Sutterer C., Becker M., Ozersky P.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
-!- FUNCTION: Transcriptional repressor that plays an important role
    in the specification and differentiation of lung epithelium. May
    play important roles in developing neural, gastrointestinal and
```

```
CC   cardiovascular tissues. Involved in neural mechanisms mediating
CC   the development of speech and language.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/I (shown here), 2/II and
CC   3/III/IV; are produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in embryonic and
CC   adult lung.
CC -!- DISEASE: Defects in FOXP2 are the cause of an autosomal dominant
CC   speech and language disorder with orofacial dyspraxia, also known
CC   as speech-language disorder 1 (SPCHI). Affected individuals have a
CC   severe impairment in the selection and sequencing of fine
CC   orofacial movements, which are necessary for articulation. They
CC   also show deficits in several facets of language processing (such
CC   as the ability to break up words into their constituent phoneme)
CC   and grammatical skills.
CC -!- DISEASE: Disruption of FOXP2 by a chromosomal translocation
CC   t(5;7)(q22;q31.2) is the cause of severe speech and language
CC   impairment.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C2H2-TYPE ZINC FINGER.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF337817; AAL10762.1; -.
DR   EMBL; U80741; AAB91439.1; -.
DR   EMBL; AC003992; AAB96326.1; -.
DR   MIM; 602081; -.
DR   MIM; 605317; -.
DR   PROSITE; PS00657; FORK_HEAD_1; FALSE_NEG.
DR   PROSITE; PS00658; FORK_HEAD_2; FALSE_NEG.
DR   PROSITE; PS00039; FORK_HEAD_3; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Chromosomal translocation; Disease mutation;
KW   Alternative splicing.
FT   DNA_BIND      504    594       FORK-HEAD.
FT   ZN_FING       346    371       C2H2-TYPE.
FT   DOMAIN         53     56       POLY-GLN.
FT   DOMAIN        123    126       POLY-GLN.
FT   DOMAIN        131    136       POLY-GLN.
FT   DOMAIN        152    191       POLY-GLN.
FT   DOMAIN        200    209       POLY-GLN.
FT   DOMAIN        223    231       POLY-GLN.
FT   VARSPLIC        1     92       MISSING (IN ISOFORM 3).
FT   VARIANT       553    553       R -> H (IN SPCHI).
FT                                  /FTId=VAR_012278.
FT   CONFLICT      134    134       Q -> H (IN REF. 2).
FT   CONFLICT      290    304       DLTTNNSSSTTSSNT -> EHFPVQGPAAVCAGL (IN
FT                                  REF. 2).
SQ   SEQUENCE  715 AA;  79919 MW;  4F9FEDB6D905I6E0 CRC64;
```

```
alignment_scores:
              Quality:        196.00        Length:        715
     Percent Similarity:       Ratio:       1.181        Gaps:        353
                                47.025        Percent Identity: 24.929        14

alignment_block:
us-09-664-641-10 x FXP2_HUMAN

Align seq 1/1    to: FXP2_HUMAN    from: 1    to: 715

850 GAAAGGGAAGGAAGGAGGAGGAGGAAGTAGAAAATGAGGAACAAGAATTCTCAGAA 899
         :::||:::          ||| ||||||||||||||||||||          :::|||||
         3 GlnGluSerAlaThrGluThrGluThrIleSerAsnSerSerMet...AsnGlnAs 18
```

```
 900 TGAGGGTAGTACAGATGAGAACTCAAGCCCCTGCCAGCTCTCAAGAAGGGT  949
        ||||||   :::::  |||:::|||:||:::|||||:::||
  18 nGlyMetSerThrLeuSerSerGlnLeuAspAlaAlaGlySerArgAspGlyA  35

 950 CTCCTTCAGTGACCAGCAGTTTCACCTAAATCCAACACTGAAAAATCT      999
        ::|||||    |||  :::|||||||::||:::|||    |||
  35 rgSerSerGlyAsp.................ThrSerSerGluValSer       45

1000 AAAGGGAATTAAATGTTGATGAATTCTTCAGATTCATCACCGGAAAAACA   1049
        ||||:::   :::||||||||:::::||:::||||:::|||
  46 ThrValGluLeuLeuHisLeuGlnLeuGlnIleAla.................  57

1050 CGAGAGAAATTTAAACTGGACCCCGCCCAAAGTCCCACACGTTACTCTGCAG 1099
        ::||| :::|||:::::||||||::::|||::|||:::||| :::
  58 ...............LeuGlnLeuAlA                            61

1100 CAAAAGCAGGCTGCCTCAGGGAAAGGAGCACCTGGGTTGATTAACTTGTGT 1149
        ::|||||  ::::|||||:::|||||:::||||::::|||:||
  61 laArgGlnLeuLeuLeuGlnGlnGlnThrSerGlyLeu               73

1150 GCCAATGTCCCACCCGTGTCCCAGGTAACATTTGTCCCCTGAGGTCCGGGG 1199
        ::||||   :::|||||:::|||||:::||||::||:::|||
  73 ..................                                    73

1200 TAATTTAAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAA   1248
        :::||||:::|||||::::|||||:::||||:::
  74 .............LysSerProLysSerSerAspProLysGlnArgP      84

1249 .............ATGAGTAGCTACCTGGAGTCCAGCTGTACGGACA      1281
                    ||||||:::|||||:::||||:::||||:::|||
  84 rGluGlnValProValSerValAlaMetMetThrProGlnValIleThr   100

1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGTGAACCGGCCATC    1331
        ::||||:::|||||:::||||:::|||||:::||||:::
 101 ................                                      101

1332 AAATGTAGCACATATCTTACAGACTCTTTCAGCCACCTACGAAAAAATTTAG 1381
        :::|||||:::|||||:::|||||:::|||||:::||||::
 102 nGlnMetGlnGlnIleLeuGlnGlnValLeuSerProGlnValIleGlnLeu 119

1382 AACACCAGGTGAATCACAGCCAGCAGGGACATTACAAAATGCCAATGCAGTG 1431
        ::||||:::|||||:::|||||:::||||:::||||:::|||||
 119 lnAlaLeuLeuGlnGlnGlnGln                               128

1432 CTGTTTAGCCAA                                          1460
        |||
 129 MetLeuGlnGlnGlnLeuGluPheTyrLysGlnGlnGlu              145

1461 CACACAC......ATGCTACAGCAGCACCAGGCCCAGCACCAGCAGC       1504
        :||:::          ||||||:::|||||:::||||:::|||||
 145 nLeuHisLeuGlnLeuLeuLeuGlnGlnGlnGlnGlnGlnGlnGln       162

1505 AGCACCACCGGTTTTACACCTTCAGCCCCAGATAATGCAGCTCCAG       1554
        ::|||||:::|||||:::||||:::|||||:::|||||::
 162 lnGlnGlnGln......GlnGlnGlnGlnGlnGlnGlnGlnGln          174

1555 CAGCAGCAGCAGCAGAATCTCCAGCAACCTTACCCCCAGCAGCCGCC      1604
        |||||:::|||||:::|||||:::||||:::|||||:::
 175 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl      191

1605 GCATCCA                                               1642
        |||
 191 nHisProGlyLysGlnAlaLysSerLeuLeuSerLeuArgGlnGlyLe     208

1643 CACCACCAATCGCCTCAGCAG                                 1662
        ::
 208 lnGlnLeuAlaLaGlnLeuVaLPheGlnGlnLeuGlnLeuLeuGlnMet    224

1663 CATCAGCTTTTGGACATGATCAGCAGTGGAGTTCAGAGAAGAAGGCTT     1712
        ::|||||:::|||||:::|||||:::|||||:::||||::
 225 GlnGlnLeuGlnGlnHisLeuSerLeuGlnLeuArgGlnGlyLe         241

1713 CTTATTG.......GGATGTGTCTTTGCAATTGCGGATATCCAG         1750
        ::                :::|||||:::|||||:::
```

seq_name: SwissProt_40:MCM1_YEAST

```
 241 ullSerIleProProGlyValnAlaAlaLeuProvalGlnSerLeuProG  258
        ::::::   ::  ||||  :::
1751 AGCAG...ATGGCTGATAAGCACAACTGCCTGGCCACCTGGAAAAGGATAATC 1797
        ::  |||||:::|||||:::|||||:::||||  |||:||
 258 lnAlaGlyLeuSerProAlaGluLeuGlnIleLeuTrpIysGluValThr   274

1798 CAGGCACAT 1806
        :::|||
 275 GlyValHis 277
```

seq_documentation_block:

```
ID   MCM1_YEAST      STANDARD;      PRT;    286 AA.
AC   P11746;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Pheromone receptor transcription factor (GRM/PRTF protein).
GN   MCM1 OR FUN80 OR YMR043W OR YM9532.08.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9024735; PubMed=2159934;
RA   Ammerer G.;
RT   "Identification, purification, and cloning of a polypeptide
RT   (PRTF/GRM) that binds to mating-specific promoter elements in
RT   yeast.";
RL   Genes Dev. 4:299-312(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88030692; PubMed=3311883;
RA   Dubois E., Bercy J., Descamps F., Messenguy F.;
RT   "Characterization of two new genes essential for vegetative growth in
RT   Saccharomyces cerevisiae: nucleotide sequence determination and
RT   chromosome mapping.";
RL   Gene 55:265-275(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89141759; PubMed=3066908;
RA   Passmore S., Maine G.T., Elble R., Christ C., Tye B.K.;
RT   "Saccharomyces cerevisiae protein involved in plasmid maintenance is
RT   necessary for mating of MAT alpha cells.";
RL   J. Mol. Biol. 204:593-606(1988).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.25 ANGSTROMS) OF COMPLEX WITH MAT2.
RX   MEDLINE=98140702; PubMed=9490409;
RA   Tan S., Richmond T.J.;
RT   "Crystal structure of the yeast MATalpha2/MCM1/DNA ternary complex.";
RL   Nature 391:660-666(1998).
CC   -!- FUNCTION: INTERACTS WITH THE ALPHA-2 REPRESSOR OR WITH THE
CC       ALPHA-1 ACTIVATOR THEREBY REGULATING THE EXPRESSION OF MATING-
CC       TYPE-SPECIFIC GENES.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE MADS DOMAIN FAMILY OF TRANSCRIPTION
CC       FACTORS.
```

```
CC   -------------------------------------
DR   EMBL; X52453; CAA36691.1; -.
DR   EMBL; X14187; CAA31289.1; -.
DR   EMBL; M17511; AAA34609.1; -.
DR   EMBL; Z48502; CAA88409.1; -.
DR   PIR; A34599; A34599.
DR   PDB; 1MNM; 18-MAR-98.
DR   TRANSFAC; T00500; -.
DR   TRANSFAC; T00501; -.
DR   SGD; S000004646; MCM1.
DR   InterPro; IPR002100; MADS-box.
DR   Pfam; PF00319; SRF-TF; 1.
DR   PRINTS; PR00404; MADSDOMAIN.
DR   SMART; SM00432; MADS; 1.
DR   PROSITE; PS00350; MADS_BOX_1; 1.
DR   PROSITE; PS50066; MADS_BOX_2; 1.
KW   Transcription regulation; DNA-binding; Nuclear protein; 3D-structure.
FT   DOMAIN       18     72       MADS.
FT   DOMAIN       98    120       ASP/GLU-RICH (ACIDIC).
FT   CONFLICT      9      9       P -> S (IN REF. 3).
FT   CONFLICT     37     37       S -> F (IN REF. 3).
FT   CONFLICT    156    157       GA -> AR (IN REF. 3).
FT   CONFLICT    158    286       MISSING (IN REF. 3).
SQ   SEQUENCE   286 AA;  32802 MW;  FD75B4BF5A9E83B CRC64;
```

```
alignment_scores:
                Quality:  192.00
                  Ratio:    1.324
     Percent Similarity:   41.429     Percent Identity: 23.143

alignment_block:
US-09-664-641-10 x MCM1_YEAST

Align seq 1/1   to: MCM1_YEAST   from: 1   to: 286
```

```
                                                        Length:     350
                                                          Gaps:      12

 667 AATAAGAAAATGCACGCATTTGATTGTTCCAGCCAAAGGGGGGGGAAATA 716
     |||||                           ||   |  |||||||
  28 AsnLysThrArgArgHisValThrPheSerLysArgLysHisGly...... 42

 717 CGAATGTGCTTTAAAGGAGCA.....AGTATTAAAATTGTGACTCCTGACT 763
     ::|||:||                   |                :||||
  43 .......IleMetLysLysAlaPheGluLeuSerValLeuThrClyThrG 57

 764 GGGTTCTCGGATTCGGTATCAGAGAAAACAAAAAGGACGGAAGCATTTTAT 813
     |||||       ||||                   || ||  :::|||
  57 InValLeuLeuLeuValValSerGluThrGlyLeuValTyrTyrThrPheSer 73

 814 CATTCTCCTCTG................ATTATTTATTCAGACAGACAA 843
     |||||                        ::|:||:||:::|||
  74 ThrProLysPheGluProIleValThrGlnGlnGlyArgAsnLeuIl 90

 844 ........................GAAGGGAAGAGGAAGAGGAGGAGGAAG 871
                             |||  ::|||  || || || ||  ||
  90 cGlnAlaCysLeuAsnAlaProAspGluGluGluGluGluGluGluGlua 107

 872 TAGAAAATAGGAACAAGAGATTCTCAGAAATGAGGGTAGTACACAGATGAGAAG 921
     :||||||||::|:::||       |||||:||||||||:||::||::::
 107 spGlyAspAspAspAspAspAspAspAspAspGlyAsnAspMetGlnArg 123

 922 TCAAGCCCTGCCCAGCTCTCAAGAAAGGGTCTCCTTCAGGTGTCGACCAGAGTT 971
     ::|||:|||  :|::|||          |||| ||
 124 GlnGlnProGlnGlnGlnGln........................ 130

 972 TTCACCTAAATCCAACACTGAAAAATCTAAGGGGAATTAATGTTTGATG 1021
     ||:|||:||||||||||||||||||||||||||:::::||||||
 130 ..............................................GAATTAATGTTTGATG 1021

1022 ATTCTTCAGATTCATCACCGGAAAAAACAGAGAGAAAATTTAAACTGGACC 1071
     |||||:||:||:||||:||:|||||||||||||||||||| |||||
 131 .......................................ProGlyGlnGlyGlnValLeuAsnAlaHis 141

1072 CCGGCGCAAGTCCCACAGTTAGCTCCGCAGCAAAAACGCAGGCGCTCCTCAGG 1121
     |||||  ||||    ||:||||:||||||||||||||:||||||||||
 141 .......................ProGlnGlnGlnGlnGlnValLeuAsnAlaHis 141
```

```
DR   EMBL; X92517; CAA63278.1; -.
DR   EMBL; Z71437; CAA96048.1; -.
DR   HSSP; P05132; 1BKX.
DR   SGD; S005105; YNL161W.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00433; pkinase_C; 1.
DR   SMART; SM00133; S_TK_X; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Hypothetical protein; Transferase; Serine/threonine-protein kinase;
KW   ATP-binding.
FT   DOMAIN       23     34       POLY-GLN.
FT   DOMAIN      160    164       POLY-GLN.
FT   DOMAIN      214    232       POLY-SER.
FT   DOMAIN      235    250       POLY-GLN.
FT   DOMAIN      352    672       POLY-GLN.
FT   DOMAIN      358    366       PROTEIN KINASE.
FT   NP_BIND     381    381       ATP (BY SIMILARITY).
FT   BINDING     381    381       ATP (BY SIMILARITY).
FT   ACT_SITE    475    475       BY SIMILARITY.
SQ   SEQUENCE   756 AA;  86946 MW;  87EBCD2C3C96EE11 CRC64;
```

alignment_scores:

Quality:	189.50	
Ratio:	0.562	
Percent Similarity:	41.968	Percent Identity: 21.669

alignment_block:
US-09-664-641-10 x KNQ1_YEAST

Align seg 1/1 to: KNQ1_YEAST   from: 1 to: 756

Length:	803
Gaps:	37

```
835 CAAGAGGAAGCAAGGAGGAAGAGGACGAAGAGCGAAGTAGAAAATGAGGA 884
    ::::::::::::: ::::::::::: ::::::::::::::::: |||||
 23 GlnGlnGlnAspGlnGlnHisGlnGlnGlnGlnIleTyrAlaAsnGluMe 39

885 ACAAGATTCTACGAATCAGGGTAGTACAGATCGAAGGTCAAGCCCCTGCCA 934
    :::                                   ::: |||
 39 tAsnProTyrGln............................GlnIleProArgP 48

935 GCTCTCACGAAAGGGTCTCCTTCAGGTGACCAGCACGTTTCACCTAAATCC 984
    |||||||||||                     :::::|||||||||||||
 48 roProAlaArgGly.........................PheSer.....Ser 55

985 AACACTGAAAAATCTAAAGGGGAAAATAATGTTGTTGATGATTCTTCAGATTC 1034
    |||  :::::::::::::::::   :::|||  :::|||||||||||||
 56 AsnThrMetLysGluGlnSer...........HisGlnSerLeuGlnGl 69

1035 ATCACCGGAAAAACACGGAGAGCAAAATTTAAACTGGACCCGGAGCTCC 1084
     |||  ::::::::::::::::::::::|||:::||
 69 uHisLeuGlnArgGlnThrGlyAsnLeuGlySerGlyPheThrAspValP 86

1085 CACAGTTAGCTCAGGAAAAACGGGCTGCCTCAGGAAAGGAGCCTGGG 1134
     |||  |||  ::::::::::::::::::::::
 86 roAlaLeu.................................... 88

1135 TTGATTAACTTGTTGCGTGCCAATGTCCACCCGTTCCCAGGTAACATTTTGCC 1184
     |||  |||                                  ::::
 89 .....................................AsnTyrProAlaThrPro 95
```

```
1185 CCCTGAGGTCCGGGGTTGAATTTAATGGCTGCTGGGACAAAACCTCCAAAGTT 1234
     ::::::::::: |||                        ::::::::||| 
 95 oPro.........HisAsnAsnThrAlaAlaSerAsnGln.......... 105

1235 CTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCGTACGGACACTG 1284
     ::::::::::::::: |||                  ::::||
106 .........MetIleAsnThrProProSerMetGlyGlyLeu 117

1285 AGGAATATTACTAATAATGCTGAC........ATTCAGCAGATCGAACCGGCC 1328
     :::|||||||||||||||||             :::|||:::|||
118 TyrArgHisAsnAsnSerGlnSerMetValGlnAsnGlyAsnGlySe 134

1329 ATCAAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAATT 1378
     :::::::::::::::|||||||               :::|||
134 rGlyAsnAlaGln......LeuProGlnLeuSerProGlyGlnTyrSer1 149

1379 TAGAACACAGGTGAATCACCAGCCAGCCAGGGACATACAAATGCCAATGCA 1428
     |||||||||||||||||:::||                  :::|||
149 leGluSerGluTyrAsnGlnAsnLeuAsnGlySerSerSerSerPro 165

1429 GTGGTGTTTAGCCAAGTG.............................. 1446
     |||:::|||
166 PheHisGlnProGlnThrLeuArgSerAsnGlySerSerGlyIle 182

1447 AAACTGACTCCAGAAGACACACAGCTAC.............. 1474
     :::::::|||||||                   :::||| 
182 uArgSerValLysSerPheGlnArgLeuGlnGlnGlnGluAsnValG 199

1475 AGCAGCAGCAACAGCAGCCCAGCAGCCAGCAGCAGCACCCGGTTTTACAC 1524
     |||||||||||||||:::|||               :::|||
199 lnValGlnValGlnLeuLeuSerGlnAlaGlnGlnGlnAsn......Ser 212

1525 CTTCAGCCCCAGATAATGGAGTCCAGCAGCAGCAGCAGCAGCAGACAT 1574
     |||       |||:::|||                 ::::|||
213 ArgGlnGlnGlnLeuGlnTyrGlnGlnGlnGlnGlnGln 229

1575 CTCTCAGCCAACCTTACCCCCAGCAGCGCGGCCGCCGCATCCATTTTCACAGCTAC 1624
     :::::::||||| |||            ::::||||||
229 nGlnGlnGlnHisMetGlnIleGlnGln........GlnGlnGlnG 242

1625 AGCAGCAGCAACAGCAGCCCACCCACCAT CGGCCTCAGCAGCATCAGCTTTTT 1674
     |||||||||||||||||||||               :::|||
242 lnGlnGlnGlnGlnGlnGlnSerGlnSerProValGln........ 254

1675 GGACACATCCAGCAGTGGGAGATTCCAGGAAGAGAAGGGTTCTTATTGGGATG 1724
     |||       |||                     :::|||
255 ............SerGlyPheAsnSerGly.... 260

1725 TGTGTTTGCAATTCGGATTATCCAGAGAATGTCTGATAAGCAAACTGC 1774
     ::: ::: |||                      :::|||
261 .............ThrIleSerAsnTyrMetTyrPheGluArgArgProAspLeuL 275

1775 TGGCCACCTGGAAAAGGATAATTCCAGGCACATGGCGGCCACTGTTGACCCC 1824
     |||||||||||||||||                  :::|||
275 euThrLys............GlyThrGlnAsp... 281

1825 ACCTT:ACGAGTCCATCCACCCACCCTTCCTGTGTGACAGTCAAGTCAGCAG 1874
     ::: |||                               :::|||
282 ......LysAlaAlaValLysLeuLysIleGluAsnPheTyrGlnSe 296

1875 CGGGTATGCACAGCCAATAAGAGAAGAAAGAGAAGATGTT........ 1914
     :::::::::::::::||||||||                 :::||| 
296 rSerValLysTyrAlaIleGluArgAsnGluArgArgValGluLeuGlu 313

1915 ........ACTGCACACTGGTTAAACACAGTCTTAAAAAAAAAAAAATG 1956
     :::                              :::|||
313 hrGluLeuThrSerHisAsnTrpSerGluLeuArgGlyLysSerArgGlnLeu 329

1957 GTACCCGCCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCACCAGGAGG 2006
     |||                                    :::|||
330 ...........AsnTyrProAlaThrPro.............. 330
```

  ....SerSerLeuGl 333

2007 AAAGCCATGTTCACAGAGATATTATTTCTGTGACTGGATTTGTTGATAGTG 2056
     ||||||
333 ylLysLysGluSerGln.......................... 343
                                              PheLeuArgLeuA

2057 ACAGAGATCACCTAAAAATAATGCCTTATTTGGCAGGTGCCAAATATACG 2106
     :::|||
343 rqArgThrArgLeuSerLeuGluAspPheHisThr...ValLysValIle 358

2107 GCTTATCTATGCCCAGCACAGTCCTCATCTGTAAAGACACCAACTGC 2156
     |||
359 GlyLysGlyAlaPheGlyGluValArgLeuValGlnLysLysLysLysAr 375

2157 T........................TTAAAGTATGAAAAAGGCAAAAGAGT 2182
     :::
375 ylLysIleTyrAlaMetLysThrLeuLeuLysSerGluMetTyrLysLysA 392

2183 GGAGGATACCCTGTTCACACGCCCAGTCGCTTGGCACATTCTTCTTGGA 2232
     |||||
392 spGlnLeuAlaHisValLysValLysLysAlaGlu.......... 406

2233 AAC............... TTTGAGGCACTGAGGCAGAATTCA 2258
     ||||||
407 SerAspSerProGlyTrpValValSerLeuThrTyrTyrSerPheGlnAspAlaGl 423

2259 GTATAGTCCGCTACACGGCATTCAGTCTGTCCAGGATCCATTGCCCCTACCC 2308
     :::
423 nTyr.......... LeuTyrLeuIleMetGluPheLeuProGlyG 435

2309 AGGATTTAGTTTTTAAATCTTTTAGATGCTTGGAGA............. 2343
     |||
435 lysAspLeu...MetThrMetLeuIleArgTrpGlnLeuPheThrGluAsp 450

2344 CTTTCCCTTAAAAGTGCTCTGCAGAGTTGTTGATGACTATAAAGACTACCTCC 2393
     |||
451 ValThrArgPheTyrMetAlaGluCysIleLeuLeuAlaIleGluHisIleHi 467

2394 CAAACTG.......................... AAACAGAATGAAGTAGCTA 2419
     ||||||
467 sLysLeuGlyPheIleHisArgAspIleLysProAsnIleLeuIleA 484

2420 ATGTCACCCTTCTTCCCAAAGACCAGAATTGAAGACGTACCACCTCCC 2469
     :::
484 spIle.......... ArgGlyHisIle............... 489

2470 ACTAAAAAGCTAACTCCAGAATTGGTCGCTTTCACTCGGATT 2519
     |||
490 ....LysLeuSerAspPheGlyLeuSerThrGlyPh 500

2520 CGAGGCCGTCCAGGTTCAACAGTATATTAAGAAGCTCTAC.......... 2559
     |||||||
500 cHisLysThrHisAspSerAsnTyrTyrLysGlyLeuGluGlnAsp 517

2560 .... ATTCTTGGTCGGAGGTTGCC 2580
     :::
517 .... LysLeuSerAspPheGlyLeuSerProGlyThrTyrAsnAlaAsnThrThr 533

2581 CAGTCTGCCACAGAGTCCACACACTTCATTGCAGCAAAGTGACTCGGACAC 2630
     |||||||
533 luAlaThrAsnGlyIleSerLysProGlyThrTyrAsnAlaAsnThrThr 533

2631 CGTG............................. AAGTTCCTGA 2644
     ||||
550 rMetSerAsnArgGlnGlnIleGlnThrThrArgLysSerArgArgLeuM 567

2645 CGGCGATTTCTCTGTGAAGCACATACTGACGCCAGACTGGCTG...GAA 2691
     |||||||:::|||
567 etAlaTyrSerThrValGly.......... ThrProAspTyrIleAlaPro 580

2692 GAATCGCTTCAGGTGTCAGAAGTTCATTGATGACGAACTACATTCTCCG 2741
     ||| |||
581 GluIlePheLeuTyrGlnGlyTyrGlyAsnGlu.......... 591

2742 AGATGCTGGAGGCACGAAGTACTTTCTCTTTCAGCTTGGAGAATCCTTAA 2791
                                                   591

591 591

2792 AACGGGGCACACGTTTCTCCACTCTTTAAGGCCAAAATAATTTTTACATCACA 2841

591 591

2842 CCTGGAATCTGCCCAAGTCTTCCACTATGAAGGCAATCCTGTAGAGTGTGC 2891
     |||
592 ....CysAspIleTrpSerLeuGlyAlaIleMetTyrGluCysIle 605

2892 AGGAGGA................ AAGGTGTATCCAAGCAGCCA........ 2929
     |||
605 ulleGlyTrpProProPheCysSerGluThrProGlnGlnThrTyrArgGl 622

2930 AGCTCATTGAGGCAGGCAAGAACTCGAGTTGTCGGAAATAATTTTAATA 2979
     ||:::|||||::: ||:::||| |||||
622 ystleMetAsnPheGlnIleThrLeuGlnPheProAspAspIleHisIle 638

2980 TCCTGTGAA 2988
     ||| |||
639 SerTyrGlu 641

639 SerTyrGlu 641

seq_name: SwissProt_40:SSN6_YEAST

seq_documentation_block:
ID   SSN6_YEAST   STANDARD;   PRT;   966 AA.
AC   P14922;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   Glucose repression mediator protein.
GN   SSN6 OR CYC8 OR YBR112C OR YBR0908.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89211964; PubMed=2854095;
RA   Trumbly R.J.;
RT   "Cloning and characterization of the CYC8 gene mediating glucose
RT   repression in yeast.";
RL   Gene 73:97-111(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88065502; PubMed=3316983;
RA   Schultz J., Carlson M.;
RT   "Molecular analysis of SSN6, a gene functionally related to the SNF1
RT   protein kinase of Saccharomyces cerevisiae.";
RL   Mol. Cell. Biol. 7:3637-3645(1987).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=92327848; PubMed=1626431;
RA   Mannhaupt G., Stucka R., Ehle S., Vetter I., Feldmann H.;
RT   "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the
RT   excision repair gene RAD16 located in this region belongs to a novel
RT   group of double-finger proteins.";
RL   Yeast 8:397-408(1992).
RN   [4]
RP   TPR REPEATS.
RX   MEDLINE=9012463P; PubMed=2404612;
RA   Sikorski R.S., Boguski M.S., Goebl M., Hieter P.A.;
RT   "A repeating amino acid motif in CDC23 defines a family of proteins
RT   and a new relationship among genes required for mitosis and RNA
RT   synthesis.";
RL   Cell 60:307-317(1990).
-!- FUNCTION: IT IS INVOLVED IN REPRESSION BY A1 ALPHA2 AND ALPHA2 AND
CC   IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS
CC   PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT
CC   DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.
-!- SUBCELLULAR LOCATION: Nuclear.

```
CC   -!- SIMILARITY: CONTAINS 10 TPR REPEATS.
CC   -!- SIMILARITY: TO YEAST GAL1 AND CCR4.
CC
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M23440; AAA34545.1; -.
DR   EMBL; M17826; AAA35103.1; -.
DR   EMBL; X66247; CAA46973.1; -.
DR   EMBL; X78993; CAA55615.1; -.
DR   EMBL; Z35981; CAA85069.1; -.
DR   PIR; S23365; S23365.
DR   SGD; S0000316; CYC8.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF00515; TPR; 10.
DR   SMART; SM00028; TPR; 9.
KW   Transcription regulation; Repressor; Repeat; TPR repeat;
KW   Nuclear protein.
FT   DOMAIN       15     30       POLY-GLN.
FT   REPEAT       46     79       TPR 1.
FT   REPEAT       80    113       TPR 2.
FT   REPEAT      114    147       TPR 3.
FT   REPEAT      150    183       TPR 4.
FT   REPEAT      187    220       TPR 5.
FT   REPEAT      224    257       TPR 6.
FT   REPEAT      258    291       TPR 7.
FT   REPEAT      296    329       TPR 8.
FT   REPEAT      330    363       TPR 9.
FT   REPEAT      364    398       TPR 10.
FT   DOMAIN      493    556       30 X 2 AA TANDEM REPEATS OF Q-A.
FT   DOMAIN      557    587       POLY-GLN.
FT   CONFLICT    547    547       K -> Q (IN REF. 3).
SQ   SEQUENCE   966 AA;  107202 MW;  84B509CF3208C5C0 CRC64;
```

```
alignment_scores:
            Quality:   186.00
              Ratio:     0.805
Percent Similarity:    44.509

alignment_block:
US-09-664-641-10:
Align seg 1/1  to: SSN6_YEAST

Align seg 1/1  to: SSN6_YEAST        from: 1   to: 966

                               Length:      519
                                 Gaps:       24
Percent Similarity:  44.509   Percent Identity:   23.314
```

```
1087 CAGTTAGCTGGAGCAAAACGCAGGCTGCCCCAGGAAAGAGCCTGGGTT 1136
     ::::||||  |||        || |||:||||:::|| |||||||
 404 ArgLeuGluAlaLeuThrLysGlnLeu............GluAsnProGlyAs 417

1137 GATTAACTTGTGT..........GCCAATGTCCACGGTCCCCAGGTA 1174
     |||||||                 :::|||:::  |||||:|:|||
 417 nIleAsnLysSerAsnGlyAlaProThrAsnAlaSerProAlaProProp 434

1175 ACATT..............TTGCCCCCTGAGGTCCGGGGTGTAATTTAATG 1209
                       ::||  ||| ||:|| |||||| ||||:|||
 434 roValIleLeuGlnProThrLeuGlnProAsnAspGlnGlyAsnProLeu 450

1210 GCTGCTGGACAAAACCTCCAAGTTCTGAA..AGATCAGAAAATGATAGC 1256
     ::|                 ||||||:|:::||: :|||||||||
 451 AsnThrArgIleSerAlaGlnSerAlaAsnAlaThrAlaSerMetValGl 467

1257 TACCTGGAGTGCAGCGTGTACCGACACTGAGGAATATTACTAATAATGCTG 1306
     |||||||                   ||||||| |||:|||:|||||| |||
 467 nIleAsnLysSerAsnGlyAlaProThrAsnAlaSerProAlaProProp 475

1307 ACAATTCAGCAGGATGAACCGGCCATCAAAATGTAGCACATATCTTACAGACT 1356
     :::|||   |||  |||||| |||  ||| |||
 476 ..........ProIleAsnSerSerAlaThrMetTyrSer 485
```

```
1357 CTTTCAGCACCTACCGAAAAATTTAGAACACAGCCAGCGAATCACAGCCACCA 1406
     :::|||||:  :::||::|  |||:|:::: |||:|||| ||||:::| ||
 486 AsnGlyAla...SerProGlnLeuGluAsnAlaGlnAlaGlnAlaGl 501

1407 GGGACAATACAAAATCCCAATGCACTGCAGTGCTGTGTTAGCCAAGTGAAAGTGCACTC 1456
     |||:||| |||||||:::::||||::::: ||||||:::|||:||||||:::|
 501 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnA 518

1457 CAGAGACACAC...ATGCTACAGCAGCAGCAGCGCCCAGCAGCAGCAG 1503
     ::::||||| :::  ||| ||||| ||||||||||||| |||
 518 laGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 534

1504 CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG.............. 1536
     ||| |||   |||       |||||||
 535 GlnAlaHisAlaGlnAlaGlnAlaGlnAlaGlnAlaLysAlaGlnAlaGl 551

1537 ....CAGATAATGCAGTCCAGCAGCAGCAGCAGCAGCAGATCTCAGC 1582
     :::  |||| |||||||||||||||||||||||||| :::||| |||
 551 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnG 568

1583 AACCTTACCCCCACCAGCGCCGCCATCCATTTTCACAGCAGCAGCACAGCAG 1632
     |||||||   ||||||||      :::|||||||||||||||||||||
 568 lnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 584

1633 CAACAGCAGCCACCACCCATCGCCTCAGCCAGCATCAGCTTTTTGGACANCA 1682
     |||||||||||||||||||||||||||||||||| :::||||||:||::
 585 GlnGlnGlnLeuGlnLeuGlnProLeuProAlaGln...GlnLeuGlnLysGl 600

1683 TCCAGCCAGTGGAGATT........CCAGAAGAGGCTTCTTATTCGGATGTG 1726
     |||||| |||||| :::        || ||||||||||||| |||||||:|
 600 yValSerValGlnMetLeuAsnProGlnProGlnGly........... 611

1727 TGTTTGCAATTGCGGATTATCCAGAGCAGAGTGCTGATAAGCAACTGCTG 1776
     ||||||||||||||||||||||||||||||||||||||||||||||
 612 ..............GlnProTyr 614

1777 GCCACCTGGAAAAGGATAAATCCAGGCACACATGCGCGGCACTGTTGACCCCAC 1826
     :::|||||||||||||||||::::::||||||||||||||:::|||
 615 IleThrGlnProThrValIleGlnAlaHis............GlnProPh 629

1827 CTTCACCAGTCGATGCAGCAGCACCTTCTTCTGTGAGAGTCAAGTCAGCACG 1876
     ||||||||| :::||||||||||||||:::|||||:::|||||||||
 629 eSerThrGlnAlaMetGluHisProGln...SerSerGlnLeuProProG 645

1877 CGTATGCACAGGGCAATAAAGAGAAAGAGAAGAGATGGTTACTGCACACTGG 1926
     |||  |||||||||||||||||||||||||||||||||||||||||
 645 lnGlnGlnGlnLeu.................... 649

1927 TTAAACACAGTCTTAAAAAAAAAAAATGGTTACCGCCGCCACCCGACGCCCT 1976
     :::|||||||||||||||||||||||||||::::::
 650 ...........GlnSerValGlnHisHisProGlnHisLeuGl 659

1977 TCACTTCCCAGTGCCTTCCCACCAGGAGGAAAAGCCCATGTTCACACGATA 2026
     |:::||| :::|||||||||||||||||||||||||||||||
 659 nGlyGlnProGlnAlaGlnAlaPro.....GlnProLeuIleGlnHisA 674

2027 TTATTTCTGTGACTGGATTGTTGATAAGTGACGAGAGATGACCTTAAAATTA 2076
     ::::|||||||||||||||||||||||||||||||||||||||
 674 snVal..............GluGlnAsnValLeuProGln 682

2077 ATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCCATACGTGACCGACAA 2126
     |||||||||||||||||::::::
 683 LysArgTyrMetGluGlyAlaIleHisThr.LeuValAspAlaAlaVals 699

2127 CACAGTCCTCATCTGTAAAGAACCAAACTGGTTTAAAGTATGAAAAAGCCA 2176
     :::|||||:::  ||||||||||||||||||||||||||||
 699 erSerSerThrHisThrGluAsn...........AsnThrLysSerPro 711

2177 AAGAGTGGAGGGATAACCTGTGTCAACGCCCAGTGGCTTGGCGACANTTCTT 2226
     :::|||||||||||||||||||||||||||||||||||||||
 712 ArgGlnProThrHisAlaIleProThrGlnAla................ 722
```

seq_name: SwissProt_40:GLTA_WHEAT

```
ID   GLTA_WHEAT     STANDARD;      PRT;    356 AA.
AC   P10385;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutenin, low molecular weight subunit precursor.
OS   Triticum aestivum (Wheat).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Triticeae; Triticum.
OX   NCBI_TaxID=4565;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. YAMHILL;
RX   MEDLINE=89083577; PubMed=3205747;
RA   Pitts E.G., Rafalski J.A., Hedgcoth C.;
RT   "Nucleotide sequence and encoded amino acid sequence of a genomic
RT   gene region for a low molecular weight glutenin.";
RL   Nucleic Acids Res. 16:11376-11376(1988).
CC   -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
CC       PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
CC       VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
CC   -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
CC   -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
CC       GROUP 1 CHROMOSOMES OF WHEAT.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X07747; CAA30570.1; -.
DR   PIR; S01992; S01992.
DR   InterPro; IPR003612; AAI.
DR   InterPro; IPR001768; Cereal_tryp_amyl_inh.
DR   InterPro; IPR001954; Glia_glutenin.
DR   Pfam; PF00234; tryp_alpha_amyl; 1.
DR   PRINTS; PR00208; GLIADGLUTEN.
DR   SMART; SM00499; AAI; 1.
KW   Seed storage protein; Repeat; Multigene family; Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20    356       GLUTENIN, LOW MOLECULAR WEIGHT SUBUNIT.
SQ   SEQUENCE   356 AA;  41020 MW;  AA2613FC0DD4DCA5 CRC64;
```

alignment_scores:
```
        Quality:       185.50
        Ratio:           2.576
Percent Similarity:     58.065     Percent Identity:  45.968
```

alignment_block:
US-09-664-641-10 x GLTA_WHEAT

```
Align seq 1/1  to:  GLTA_WHEAT  from: 1  to: 356

                                              Length:   124
                                              Gaps:       5
```

seq_documentation_block:

seq_name: SwissProt_40:BRD4_HUMAN

```
ID   BRD4_HUMAN     STANDARD;      PRT;   1362 AA.
AC   O60885; Q96FD3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Bromodomain-containing protein 4 (HUNK1 protein).
GN   BRD4 OR HUNK1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   French C.A., Fletcher J.A.;
RT   "Human BRD4 protein.";
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-722 FROM N.A.
```

```
RC    TISSUE=Placenta;
RA    Weber H.;
RL    Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
CC    -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC    -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC
CC    ---------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    ---------------------------------------------------------------------------
DR    EMBL; AF386649; AAL26987.1; -.
DR    EMBL; Y12059; CAA72780.1; -.
DR    InterPro; IPR001487; Bromodomain.
DR    Pfam; PF00439; bromodomain; 2.
DR    PRINTS; PR00503; BROMODOMAIN.
DR    SMART; SM00297; BROMO; 2.
DR    PROSITE; PS00633; BROMODOMAIN_1; 1.
DR    PROSITE; PS50014; BROMODOMAIN_2; 2.
KW    Bromodomain; Repeat; Nuclear protein.
FT    DOMAIN       75    147       BROMODOMAIN 1.
FT    DOMAIN      368    440       BROMODOMAIN 2.
FT    DOMAIN      535    594       LYS-RICH.
FT    DOMAIN      692    717       SER-RICH.
FT    DOMAIN      703    714       POLY-SER.
FT    DOMAIN      738    743       POLY-HIS.
FT    DOMAIN      757    761       POLY-PRO.
FT    DOMAIN      764    770       POLY-PRO.
FT    DOMAIN      771    775       POLY-GLN.
FT    DOMAIN      776    783       POLY-PRO.
FT    DOMAIN      954    964       POLY-PRO.
FT    DOMAIN      974    986       POLY-PRO.
FT    DOMAIN     1011   1014       POLY-PRO.
FT    DOMAIN     1028   1033       POLY-PRO.
FT    DOMAIN     1283   1300       POLY-GLN.
FT    DOMAIN     1301   1308       POLY-ALA.
FT    DOMAIN     1335   1338       POLY-ARG.
FT    CONFLICT    720    721       EM -> GP (IN REF. 2).
SQ    SEQUENCE   1362 AA;  152219 MW;  D52EFF1CF9960907 CRC64;
```

alignment_scores:
```
                Quality:  182.50
                Ratio:      1.037
Percent Similarity:  40.741  Percent Identity:  23.511

alignment_block:
US-09-664-641-10 x BRD4_HUMAN

Align seq 1/1 to: BRD4_HUMAN  from: 1 to: 1362       Length:   432
                                                     Gaps:      21
```

```
 658 CTAACCCTCAATAAGAAAATGCACGCATTTGATTCGTTCCAGACCCAAAGGG 707
     ||:|||| ||::||||||||||
 620 .....LeuAspIleAsnLys.......................LeuProGlyGluLysIle 630

 708 GGGGAAA.......TACGAATGTGCTTTAAAGCGAGCAAGTATTAAAATTG 751
     ||:|||| :::        ||| ||||||
 630 uGlyArgValValHisIleIleGlnSerArgGluProSerLeuLysAsnS 647

 752 TGACTCCTGAC.......                                762
     ::|||||||||
 647 erAsnProAspGlnIleGluIleAspPheGluThrLeuLysProSerThr 663

 763 ...............TCGGTTCTCGGATTCCGCTATCAGAGAGAAACCAA 794
                    ::||| :::|||||||||
 664 LeuArgGluLeuLeuArgThrValThrSerCysLeuArgLysLysArgLy 680

 795 AAAGGACGAGGAGCAATTTATCATCCTCGTCGATTAGTTATTATGAAGAGGAG 844
     |||||| ||||||||
 680 s...........................................          680
```

```
 845 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGTACAAAAATGAGGAAGAGATTCT 894
         |||::||||::::|| |:::
 681 .........ProGlnAlaGluLysValAspValIleAlaGlySerSer 693

 895 CAGAATGCAGGGT..........AGTACGATGAGGAAGTCAAGCCCTCCAGCTTC 938
     |||||||||       :::      ||| |||  ||||||||||
 694 LysMetLysGlyPheSerSerGluSerGluSerSerGluSerSe 710

 939 TCAAGAGAGGTGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAAATCCAACA 988
     |||||||     :::::||:  |||      :::||||||||||
 710 rSerSerSerAspSerGluAspSerGluThrGluMetAlaProLysSer 725

 989 CTGAAAAATCTAAAGGGGAATTAATCTTCAATGATTCTTCACATTCAICA 1038
         |||::|||||||:|||||
 726 ..............LysLysLysGlyHis.................... 730

1039 CCGGAAAAACAGGAGAGA............................AATTTAAA 1064
     :::
 731 ProGlyArgGlnLeuGlnLysLysHisHisHisHisHisHisGlnGlnMetGl 747

1065 CTGGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGACAGGCTC. 1114
     :::||||||| ||||||||||
 747 nGlnAlaProAlaProValProGlnGlnProProProGlnGlnGlnGlnP 764

1115 CT.......................CAGGGAAAGGAGCCTGGGTTGAATTAAC 1143
     ||                       ||||||||||
 764 roProProProGlnGlnGlnGlnGlnGlnPro................... 776

1144 TTGTGTCCCAATGTCCCACCCGTCCCAGGTAACATTTTTGCCCCTGAGGT 1193
         ||||||  ||| ||||                :::||| :::
 777 ...........ProProProProProSerMetProGlnGln... 788

1194 CCGGGGTA..TTAATGGCTGGTGCTGGACAAAACCTCCCAAAGTTCTGAAAGAT 1243
     :::         |||||||||
 789 ................AlaAlaProAlaMetLysSerSerPro....P 798

1244 CAGAAAATGATAGCTCACCTGGGAGTCCAGCTGTACGGACACACTCAGGAAATATT 1293
     :::|||||||||||||
 798 roProPheIleAlaThrGlnValProValLeuGlnProGlnLeuProGly 814

1294 ACTAATAATCCTGACATTCAGGAGATGAACCGGCCANTCAAAATGTAGCACA 1343
     |||  :::|||::|||||||||
 815 SerValPheAspProIleGlyHisPheThrGlnPro......lleLeuHi 829

1344 TAICTTACAGACTCTTTCAGCACTTACGAAAAAAATTTAGAACAGCAGGTGA 1393
     |::: |||    ||| |||
 829 sLeuProGlnProGlnLeuProPro.......HisLeuProGlnProPro 844

1394 ATCACACGCCAGCGCAGGAGCACATACAAATGCCAATGCACTGTTAGCCAA 1443
     ::|||||||                                :::||||||:::
 844 luHisSerThrProProHisLeuAsnGlnHisAlaValValSerProPro 860

1444 GTG...........                                   1446
     |||
 860 .............AAAGTGACTTCCA................... 1458

1447 ...............                                  1458

1459 .............GAGACACACATGCTACAGCGAGCAGCAGCAGCACCAG 1503
                  :::||| :::|||   ||||||||||||||||
 861 AlaLeuHisAsnAlaLeuProGlnGlnProSerArgProSerAsnArgA 877

1504 ...............CAGCAGCACCCGGTTTTACACCTTCAGCCCCAGCA 1538
                    |||  :::||||||||||||||
 877 aAlaAlaLeuProProLysProAlaArgProProAlaValSerProAlaL 894

1539 GATAATGCAGCTCCACAGCCAGCAGCAGCAGCAGCACGAATC.......... 1575
     |||||||||      |||||||||
 894 euThrGluThrProLeuLeuProGlnProMetAlaGlnProGlnLeuPro 910

 911 ValLeuLeuGluAspGluGluProProAlaProProLeuThrSerMetGl 927

 927 n...MetGlnLeuTyrLeuGlnGlnLeuGlnLysValGlnProProThrP 943
```

```
1576 .......................TCTCAGCAACCTTACCCC....CAG 1596
 943 roLeuLeuProSerValLysValGlnSerGlnProProProLeuPro 959
1597 CAGCCGCCGCCATTTTCACAGCAGCAGCAGCAGCAGCAGCCACC 1646
      ||||||||||| |||| ||| |||||||||||||||||
 960 ProProProHisProSerValGlnGlnGlnLeuGlnGlnGlnProProPr 976
1647 ACCATCGCCT......................CAGCAGCATCAG 1668
      |||| |||                        ||||||||||||
 976 oProProProGlnProGlnProProProGlnGlnHisGln 991
```

seq_name: SwissProt_40:INVO_PIG

seq_documentation_block:
```
ID   INVO_PIG      STANDARD;      PRT;      347 AA.
AC   P18175;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Involucrin.
GN   IVL.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90348475; PubMed=2385171;
RA   Tseng H., Green H.;
RT   "The involucrin genes of pig and dog: comparison of their segments of
RT   repeats with those of prosimians and higher primates.";
RL   Mol. Biol. Evol. 7:293-302(1990).
CC   -!- FUNCTION: INVOLUCRIN IS A KERATINOCYTE PROTEIN THAT FIRST APPEARS
CC       IN THE CELL CYTOSOL, BUT ULTIMATELY BECOMES CROSS-LINKED TO
CC       MEMBRANE PROTEINS BY TRANSGLUTAMINASE. ALL THAT RESULTS IN THE
CC       FORMATION OF AN INSOLUBLE ENVELOPE BENEATH THE PLASMA MEMBRANE.
CC   -!- TISSUE SPECIFICITY: PRESENT IN KERATINOCYTES OF EPIDERMIS AND
CC       OTHER STRATIFIED SQUAMOUS EPITHELIA.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; M34441; AAA31058.1; -.
DR   InterPro; IPR002360; Involucrin.
DR   PROSITE; PS00795; INVOLUCRIN; 1.
KW   Keratinocyte; Repeat.
SQ   SEQUENCE   347 AA;   41408 MW;   3DE579871CF1C829 CRC64;
```

alignment_scores:
```
         Quality:   182.00
         Ratio:       0.958
Percent Similarity:  43.981
```

alignment_block:
```
US-09-664-641-10 x INVO_PIG
```

Align seg 1/1   to: INVO_PIG   from: 1   to: 347
```
                                    Length:      432
                                    Gaps:         16
Percent Similarity: 43.981   Percent Identity: 22.454
```

```
 724 GCTTTAAAGCGGACGAAGCTATTATAAAATTGTGACTCCCTGACTGGGTTCTGCA 773
     ||||| |||||  ||||||: :||||||||: :||
  14 AlaLeuSerGlnGluProLeuUProValSerProPro.....AlaAs 28

 774 TTGCGTATCAGAGAGAGAAAACCAAAAAG.................. 798
     :::||:|:|:||:|
  28 pThrGlnGlnGluGlnValLysGlnAlaThrProLeuProAlaProCysG 45
```

```
 799 ......................................GACGAACCACATTT 810
     |::|||||  ||                            :::::|:||  :|||
  45 lnLysMetLeuSerGlnLeuProValGluValProLeuGluGluHisAspAlaGlu 61

 811 TATCATCCTCGTCTGATT......ATTTATGAAGAGAA....GA 845
     |||| |||| ||||||||      ::: |||:::|||    ::
  62 LysHisThrThrLeuValLysGlyValProGlnGlnCysGluProGl 78

 846 AGAGGAAGAGAGGAAGAGAGGAAGAGAGGAAGAGTAGAAAATGAGGAACAAGATTCTC 895
     ||:||||||||||||:|:||||:|::|:|:|||||||||
  78 nProGlnProGlnProGlnGlnGlnLeuGlnGlnGlnHisGlu 95

 896 AGAATGAGCGGTAGT......................ACAGATGAGAAGTCAAGCCCT 930
     |::||||||:||                          ||:|:|| |||:|:|||
  95 lnGlnGlnSerGlnValGlnGlnGluHisLeuValGlnGlnGlnHisGlnGlnGln 111

 931 GCCAGCTCTCAAGAAGGGTCTCTTCAGTGACCAGCAGTTTCAACCTAA 980
     ||:|:||||| ||| |||||:::|:||||||||||
 112 GlnGluSerGlnGluGlnLeuGlnHisValAspGlnGln 125

 981 ATCCAACACTGAAAAATCTAAAGGGGAATTAATGTTTGATGATTCTTCAG 1030
     |::|||| |||||||:||||||||:|:||:|||
 125 nGlnGlnGlnLysGlnSerGlnGluGlnLeuGlnHisValAspGlnGln 142

1031 ATTCATCACCGGAAAAACAGGAGAGAAAAATTTAAACTGGACCCCGGCCGAA 1080
     ::  ||:||||
 142 lnGlnGluSerGlnValGlnGlu 149

1081 GTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCCTCAGGGAAAGGAGCC 1130
     :::::||||||
 149 ................. 149

1131 TGGGTTGATTAACTTGTGTGCCAATGTCCACCCGTCCCAGGTAACANTTT 1180
     :::|||||||
 149 GlnGlnGluSerGlnGlu 149

1181 TGCCCCCTGAGGTCCGGGGGTAATTPAATTGGCTGCTGGACAAAACCTCCAA 1230
     :::|||||||||
 149 ................. 149

1231 AGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGTCCAGCCTGTACGGAC 1280
     |||:::|:||::|||
 150 LeuHisValGlyHisHisGlnGln 157

1281 ACTGAGGAATATTACTAATAATGGTGACATTCAGAGATGAACCGGGCAT 1330
     :::::||||:||
 158 GlnGlnGluSerGlnGlu 163

1331 CAAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAAAAAATTTA 1380
     |||||:::|:|||
 164 GlnLeuHisValAspHisHisGlnGlnGlnLeuGlnGlnGluSerGlnGln 179

1381 GAACAGCAGGTGAATCACAGCCAGCAGGGACATACAAAATGCCAAATGCAGT 1430
     ::||||:||  :::::||||:|:
 179 luGlnGlnLeuHisValAspGln 186

1431 GCIGTTAGCCAAGTGAAAAGTGACTCCAGAGACACACACANTGCTACAGCAGC 1480
     |:::|:|| :|:|:
 187 GlnGlnGluHisLeuValGlnSerGlnGlu 195

1481 AGCAGCAGCCAGCCACCAGCAGCAGCAGCAGCACCCCGGTTTACACCTTCAG 1530
     ::::|:|||:::|:|:|:|:||
 196 ................GlnGlnLeuHisValAspGlnGlnGln 204

1531 CCCCAGCAGAATAACCAGCTCCAGCAG. 1565
     ||||||||
 204 lnGlnGlnGlnGluSerGlnValGlnGlnGluGlnLeuGlnGlnGlnHisGlnGln 220

1566 GCAGCAGAGTCTCT... 1579
     ||||||||:||:
 221 GlnGlnGlnGluSerGlnValGlnGlnGluGlnLeuGlnHisValAspHisHisGlnGln 237

1580 AGCAACCTTACCCCCAGCAGCCGCGCCCCATCCANTTTCAGCAGCAG... 1626
     |:|:||:|:::|||
 237 nGlnGlnGlnGlu SerGlnValGlnGlnGluGluHisLeuValAspGlnGlnHisGlnGlnGln 254
```

```
254 lnGlnGluSerGlnGluGlnLeuHisValAspGlnGlnGlnGlu 270
    :::|||          |||        ::::||||||||||
1627 CACCAGGACAACACCAGCCACCACCACCATGGCCTCA 1658
    |||||||||:|||
271 LeuGlnValGlnGlnValGlnGlnGlnGlnGln........GlnGln 283
    |||||:::|||
1659 GCAGCATCAGCTTTTGACATGATGCCAGTGAGAGAATTCCAGAAGAAG 1708
    |||::|||       |||||||||::|||
283 nGlnGluGlnGlnGlnAspHisGlnLysAlaGluHisLeuGlnHisGlu. 299
    |||::|||           :||       :||
1709 GCTTCTTATTGGGATGTGTTTGCAATTGCGGATTATCCAGACGAGATG 1758
    |||       :::    ::: ||:   :::
300 ...........GluAlaGln 302
        |||  :::
1759 TCTGATAAGCAACTGCT.GGCCACCTGCAAAAGGATAA*CCAGGCACATG 1807
    |||:||||:::|||
303 ArgLeuGlnGlnLeuLysGlyGlnLeuGluGlnLeuGlnGluLysGlyValTyr 319
    :|||     :||  |:||||||:::|||:::
1808 GCGGCACTGTTGACCCCACCTTCACGAGTCATGACGACCACCTT 1851
    |||||:::
319 rGlnHis.LeuAspGlnGluLeuThrLysArgAspGluHisLeu 333
```

---

seq_name: SwissProt_40:KAPC_DICDI

seq_documentation_block:

```
ID   KAPC_DICDI     STANDARD;     PRT;     648 AA.
AC   P34099;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   cAMP-dependent protein kinase catalytic subunit (EC 2.7.1.37).
GN   PKAC OR PK2 OR PK3
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91323730; PubMed=1864510;
RA   Buerki E., Anjard C., Scholder J.-C., Reymond C.D.;
RT   "Isolation of two genes encoding putative protein kinases regulated
RT   during Dictyostelium discoideum development.";
RL   Gene 102:57-65(1991).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=93385090; PubMed=8373760;
RA   Anjard C., Etchebehere L., Pinaud S., Veron M., Reymond C.D.;
RT   "An unusual catalytic subunit for the cAMP-dependent protein kinase
RT   of Dictyostelium discoideum.";
RL   Biochemistry 32:9532-9538(1993).
RN   [3]
RP   CHARACTERIZATION.
RC   STRAIN=AX3;
RX   MEDLINE=93066311; PubMed=1332055;
RA   Mann S.K.O., Yonemoto W.M., Taylor S.S., Firtel R.A.;
RT   "DdPK3, which plays essential roles during Dictyostelium development,
RT   encodes the catalytic subunit of cAMP-dependent protein kinase.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:10701-10705(1992).
CC   -!- FUNCTION: ESSENTIAL FOR DIFFERENTIATION AND FRUIT MORPHOGENESIS.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC   -!- SUBUNIT: IN DICTYOSTELIUM THE HOLOENZYME IS A DIMER COMPOSED OF
CC       A REGULATORY (R) AND A CATALYTIC (C) SUBUNIT. IN THE PRESENCE OF
CC       CAMP IT DISSOCIATES INTO THE ACTIVE C SUBUNIT AND AN R MONOMER.
CC   -!- DEVELOPMENTAL STAGE: CAPK ACTIVITY IS LOW IN VEGETATIVELY GROWING
CC       AMOEBAE, INCREASES DURING DEVELOPMENT OF AGGREGATION AND REACHES
CC       A MAXIMUM AT CULMINATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
```

```
DR   EMBL; M37703; -; NOT_ANNOTATED_CDS.
DR   PIR; JQ1150; JQ1150.
     HSSP; P05132; IATP.
DR   DictyDb; DD02030; pkaC.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000961; Pkinase_C.
DR   InterPro; IPR0022290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00433; Pkinase_C; 1.
     SMART; SM00220; S_TKc; 1.
     SMART; SM00133; S_TK_X; 1.
     PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
     PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
     PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding; cAMP;
KW   Phosphorylation.
FT   DOMAIN     58     64     ASN-RICH.
FT   DOMAIN    136    223     GLN-RICH.
FT   DOMAIN    233    250     THR-RICH.
FT   DOMAIN    336    590     PROTEIN KINASE.
FT   NP_BIND   342    350     ATP (BY SIMILARITY).
FT   BINDING   365    365     ATP (BY SIMILARITY).
FT   ACT_SITE  459    459     BY SIMILARITY.
FT   MOD_RES   490    490     PHOSPHORYLATION (BY SIMILARITY).
SQ   SEQUENCE   648 AA;   74458 MW;   D0F9B3A48C58D084 CRC64;
```

alignment_scores:

```
Quality:            181.50
Ratio:              0.726
Percent Similarity: 39.936
```

Align seg 1/1 to: KAPC_DICDI from: 1 to: 648

```
US-09-664-641-10 x KAPC_DICDI       Length:   626
                                    Gaps:      21
Percent Identity: 21.246
```

alignment_block:

US-09-664-641-10 to: KAPC_DICDI from: 1 to: 648

```
1027 TCAGATCATCACCGGAAAAACAGGAGAGAGAAAATTTAAACTGAGACCCCGGC 1076
     :::||||:||||:||     |||        |||      :||    |||
   2 SerAsnSerAsnAsnSerSerSerGlyAsnHisAsnSerThrThrIl 18

1077 CGAAGTCCCACACAGTTAGCTCGCAGAAAAACGCAGGCTGCCCTCAGGGAAAGG 1126
     :::|||         ||:|||            |||        |||
  18 eAsnAsnProLys................................... 22

-127 AGCCTGGGTTGATTAACTTGTGCCCAATGTCCACCCGTCCCA........ 1170
     ::::|||:::     :|||:::|||||:::|||||
  23 ......ValAsnValTyrSerAsnIleProAsnSerThrThrTyr 35

1171 .........GGTAACATTTTGCCCCTGAGGTCCGGGGGTAA 1202
              |||   ||| |||            |:||
  36 ThrTyrGlySerGlyGlyGlyGlyThrLeuSerGlyAsnAsnThrAsnAs 52

1203 TTTAATGGCTGCTGGACAAAACCTCCAAAGTTCGAAAGATCAGAAATGA 1252
     :::        :||:::|||||:::|||||||:||:::||||:||
  52 nAsnAsnThrAsnAsnAsnAsnAsnAsnAsnAsnSerSerGlyAspA 69

1253 TAGCTACCTGGAGTCCAGCTGTACGGAGCACTGAGGAAAATATTACTAATAAT 1302
     :::|||:::||||:::        :||        :||||     |||
  69 snLysAsnHisSerProValThrSerAlaThrAspAlaArgLeuThr...Lys 84

1303 GCTGACATTCAGCAGATG.......................... 1322
     ||||||||:::|||:::                          ::
  85 MetAspIleGluGluLysTrpAspAspAsnTyrGluLysAspGluAr 101

1323 CCGGCCATCAAAATGTAGCACATATCTTACAG.......... 1353
     |||      :::||||||||||:::                 ::
 101 gGluLysSerProLeuPheHisIleLeuAlaSerAsnLeuAsnSerPheG 118
```

```
1354 .........................................ACTCTTTCAGCACCTACGAAAAT 1377
 118 lysAsnPheLysValProSerThrPheSerLeuThrProProGluProAsn 134

1378 TTAGAAACAGCGATGAATCACGCCAGCCAGGACATACAAAAATGCCAATGC 1427
 135 LysGlnGlnProGlnProGlnGln.n.................... 146

1428 AGTGCTGTTTAGCCAAGTGAAAGTGAGCTCCAGAGACACACAGTACACAGC 1477
 147 ........GlnProGlnGlnGlnProGlnGlnGlnGlnProGlnGln 160

1478 AGCAGCAGCAGCCCACAGCAGCAGCAGCAGCCGGTTTTACACTT 1527
 160 lnGlnGlnGlnGlnProGlnGlnGlnGlnProGlnGlnLeu 176

1528 CAGCCCAGCACAATAATGCAGCGCGAGCGCGAGGCAGCAGATCTC 1577
 177 GlnGlnAsnGlnGlnGlnGlnGlnGlnProGlnGlnLeuGl 193

1578 TCAGCAACCTTACCCCCAGCGCGCGCAGCCCATCCATTTCACAGCAGCAGC 1627
 193 nGlnGlnLeuGlnGlnGATTATCCAGAGCAGAGTGTCTTGATAAGACAGCTGG 210

1628 AGCAGCAAGACAGCAGCCACCACCATCGCCTCAGCAGCCATCAGCTTTTTGGA 1677
 210 lnGlnGlnLysGlnGlnGlnGlnGlnGlnHisLeuHis 226

1678 CATGATCCAGAGTGGAGAATTCCAGAGAGAGAGGCTTCTTATTGGANTGTGT 1727
 227 GlnAspGlyLeuValAsnThrPro......... 234

1728 GTTTGCAATTGGGATTATCCAGAGCAGAGTGTCTTGATAAGACAGCTGG 1777
 235 .......SerThrThrGlnThrSerT 241

1778 CCACCTGCGAAAAGGAATAATCCAGCCACACATCGGCCACTGTTCACCCCACC 1827
 241 hrThr.......ThrThrThrThrThr 248

1828 TTCACGAGTCGATCGCACCGCACCTTCTCTGTGACGTCAAGTCAGCAGCGC 1877
 249 ThrThrAsnProHisThrSerGlyLeuSerLeuHisAlaHisSerSe 265

1878 GTATGCACAGGCAATAGAGAGAAGAGAAGAGAGAGATGTGTTACTGCACTGGT 1927
 265 rTyrThr...... 267

1928 TAAACACAGTCTTAAAAAAAAAAAAATGGTACGCGCGCACCGACCCCTT 1977
 268 .......ProSerAsnValLeu 272

1978 CACTTCCCAGTGGGCCTTCCCACCAGGAGGAAAAGCCATGTTCAGCATAT 2027
 273 HisSerProThrHisPhe....GlnSerSerLeuProThrArgLeuAspth 288

2028 TATTCTGCACTGGATTTGTTGAGTAGTGACGAGAGATGACCTAAAAATAA 2077
 288 rAsnProIleThrThrProIleArgGlnGlnGln...... 299

2078 TGGCTTATTTGGAGGTGCCAAATATACGGGTTATCTATCGCGAGCAAC 2127
 299 ...... 299

2128 ACAGTCCTCATCTGTAAAGAACCAACTGGTTAAAGTAGTAGAAAAAGCCAA 2177
 299 ...... 299

2178 AGAGTGGAGGATACCCTGTGTCAACGCCCAGTGGGCCTTGGGCGACATTCTT 2226
 311 nGlnIleProProProThrValAsnSerPhePheLeuProProProValA 328

2227 .........................CTGGGAAACTTTGAGGCACTGAGGCCAGAATTCAGTAT 2262
```

```
 328 snAlaArgGluArgLeuLysGluProLysGlnIleArgValLeuGlyThr
     :::||||||  :|| ::: ||||  :::
 328 AGTCGGCTACACGGCGATTCAGTCTGCAGGATCCATTTGCCCTACCCAGCA 2312
 345 GlyThrPheGlyLysVal                      Ty 351
 2313 TTTAGTTTTTAAATCTTTTAGATGCGTTGGAGAGTTCCCTTAAAAGTGTCTG 2362
 351 rIleuHisGlnAsnThrLysAspGlyCysTyrTyrAlaMetLys  365
 2363 CAGAGTTGTTCGATCAGTATAGAGACTACCTCCCAAAACTGAAACAGAATGAA 2412
 366 ......CysLeuAsnLysAlaThrValValGlnLeuLysGlnValGlu 379
 2413 GTAGCTAAGTCCAGCCTTCTTCCCAAAGAGCCAGAAATTGAAGACGTACC 2462
 380 HisLeuAsnSerGlyLySerIle............ 387
 2463 ACCTCCCACTAAAAAGCTAACTCAGAATTGACCCCTTTTGTG......C 2506
 388 ......LeuSerSerIleHisHisProPheIleValAsnL 399
 2507 TTTTCACTCGGATTCGAGCCTTCAACAGTTATTATTAAGAAGCTC 2556
 399 euTyrGlnAlaPhe............GlnAspGlyLeuSerLeu 409
 2557 TACATTCTT..........GGTGGAGAGGTTGCCGGAGTCTGCACA 2591
 410 TyrLeuLeuPheGluTyrValAlaLeuGlyLeuGluVal......... 421
 2592 GAAGTGCACACACCTCATTCCAGCAAA...GTGACTCGCACCGGA 2635
 422 ....PheThrHisLeuArgLysSerMetLysPheSerAsnSerThrAlaL 437
 2636 AGTTCCTGACGGGCGATTTCTGTCCGTGAAGCAGATAGTGACGCGCAGACTGG 2685
 437 ySPheTyrAlaAla.......... 441
 2686 CTGGAAGAATGCTTCAGGTGTCAGAAGTTCATTGATGAGCAGAAACTACAT 2735
 442 ......GluIleValAlaLeuGluPheLeuHisLysGlnAsnIleVa 456
 2736 TCTCCGACGATGCGTGAGGCAGAAGTACTT 2763
 456 lTyrArgAspLeuLysProGluAsnLeu 465
```

GAL11 AND GAL4 MAY SERVE TO EXPEDITE PHOSPHORYLATION OF GAL4.

```
CC   ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------
DR   EMBL; M68870; AAA35254.1; -.
KW   Transcription regulation; Activator; Galactose metabolism.
FT   DOMAIN       75     84       POLY-GLN.
FT   DOMAIN      257    287       GLN-RICH.
FT   DOMAIN      364    411       GLN-RICH.
FT   DOMAIN      456    464       POLY-GLN.
SQ   SEQUENCE   1008 AA;  114831 MW;  436D1EBAEA17F7DB CRC64;
```

alignment_scores:

```
          Quality:    180.50
            Ratio:     0.520
Percent Similarity:   44.774        Percent Identity:  21.032
```

alignment_block:
US-09-664-641-10 x GALY_KLULA

Align seq 1/1 to: GALY_KLULA

```
693 TCCAGAGCCCAAAGGGGGGAA................................ 713
    ||||||  ||:||||||||||
  2 SerArgIleAsnGlyGlyGlyLysAlaAsnPheAsnLeuGluLysLeuLy  18

714 ....ATACCAAGTCTGCTTTAAACGCGACCAAGTATTATTAAAATTGTGACTCC. 758
    ||:::||||| :||||||||||||||||||||||| :::||||||||
 18 SLysSerAlaGluGlnPheGluSerSerLeuTyrAlaSerCysAsnSerL  35

759 ..........................TGACTGGGTTCTGGAATTGCGTATCAGAGAAAAC 791
    ||
 35 ysAspLeuTyrLeuAspAlaMetArgLysArgIleAlaIleAlaMetAspSer  51

792 CAAAAAGGACGGAAGCATTTTATCATCCTGTCTGATTATTTATTGAGAGAGG 841
    ||    :||:||||| ::::||  :::||||
 52 AlaLysArgLysSerIleGluValAlaGln.Arg................ GlnLysL  64

842 AAGAGAGCAAGAAGGAGCAAGACGAGGAGCAAGTAGAAAATCAGGAACAAGAT 891
    :::::||||||||||::::||||
 64 euAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaValGlnAlaGlnGlnGln  80

892 TCTCAGAATGAGGGTAGTACAGATAGAGGAAGTCAAGCCCTGCCAGCTCTCA 941
    :::||:||:||    ::::|||::||||||||
 81 GlnGlnGlnHisHisMetGlnGlnGlnGlySerGlyIleGlnGlnAl  97

942 AGAAGGGTTCCTTCAGGTGACCAGCAGTTTCACCTAAATCCAACACTG 991
    ::||| |||:| :::::||
 97 aAsnGlyIleProAla...................... AsnIleAsnAlaG 107

992 AAAAATCTAAAGGGGAATTAAACTTTGATTCATTCAGATTCATCACCG 1041
    |||||:
107 In................................. MetPheLeuAsnGlnAlaGlnAlaArg 117

1042 GAAAAACAGAGAGAAAAATTAAACTGGACCCGGCCGAAGTCCCACAGTT 1091
     :::::::  ::|||||||| ::::||
118 GlnGlnAlaGlnArgGlnIleArgSerThr............. Le 128

1092 AGGTGCAGCAAAAACGCAGGCTCCCTCAGGGAAAGGAGCCTGGGGTTGATTA 1141
     |::::::|| :::|||||||
128 uThrGlyGlnGlnGlnLysGlnGlnProMetProGlnProGlnPro....... 141

1142 ACTTGTGTGCCAAATGTCCCACCCGTCCCAGGTAACCATTTTGCCCCCTGAG 1191
     ||||
142 ........................... GlnProGlnGln 145
```

```
1192 GTCCGGGGGTTAATTTAATGGCTGCTGGACAAAACCTCAAAAGTTCT....... 1236
     :::  ||||||| :::
146 GlnGlnProAsnMetMetArgProGlnLeuThrLeuGlnGlnIleGlnGl 162

1237 ..........................................GAAAGATCAGGAA 1249
     |||
162 nLeuAlaAsnGluLeuLysValThrGlnIleProArgGluLeuLeuGlnL 179

1250 TGATAGCTACCTGGAGTCCAGCCTGTACGGAGCACTGGCACTGAGGAATTACTAAT 1299
     |||       :::  |||
179 ysIleProAsnLeuProProGlyIleThrThrProGlnIleThrPhe 195

1300 AATGCTGACATTCAGCAGATGAACCGGCCATCAAAATGTAGCACATATCTT 1349
     |||     ||||||||||||| :::||||||
196 TrpAla.......... GlnLysAsnArgProSerGlnAsnAspLeuIleLeu 209

1350 ACAGAACTTCTTCAGCACCTTACCGAAAAATTTA...................... 1387
     |||||||::::||       |||
209 e........................... ThrLysLysValTyrGlnLeuHisSerLeuL 220

1388 AGGTGAATCACAGCCAGCCAGCAGGGACATACAAAATGCCAATGCAGTGCTGTTT 1437
     :::|||:: ::|||:|| :::||||
220 euLeuAsnLysSerLysGlyGlnSerAspAsnAspLeuIleLeu 234

1438 AGCCAAGTGAAAGTGACTCCAGAGACACACATGCTTCAGCAGCGAGCAGCA 1487
     ||||||||||:::::||
234 ............... AsnThrSerValGlnGlnInProGlyMetGl 244

1488 GGCCCACCAGCAGGAGCACCACCAGCAGCCGGTTTTACACCTTCAGCCGCCAGC 1537
     ||||||  :::||||
244 yValGlnLnSerProAsnGlnArgMetAsnMetAsnGlnValProGlnG 261

1538 AGATAATGCAGCTCCAGCAGCAGCACCAGCAGCAGAGTCTCAGCAACCT 1587
     |||   ||||||||||||||||||||||||| :::|||
261 InGlnGlnGlnValArgGlnGlnGlnGlnGlnGln................ 275

1588 TACCCCCAGCGAGCGCGCGCATCCATTCACAGCACGAGCAGCAGCAACA 1637
     ||||||||||:::::||
276 .......... GlnGlnGlnHisLysyGl 282

1638 GCAGCCACCACCATCGCCTCAGCACCACCAGCTTTTCGGACAATCCAGG 1687
     |||||||| :::||||
282 nGlnPro.......... GlnHisGln 291

1688 CACTGGAGATTCCAGAGACAGAAGAAGGCTTCTTAAGTGGCAGTGTGTTTTGCAATT 1737
     ||||||  :::||
291 laVal...................... LeuGlyArgIleLeuAsnGlnVal 299

1738 GCGGATTATCCAGAGAGAGATGTCTGATAAGCAACTGCTGGCCACCTGGAA 1787
     |||||       :::|||
300 PheThrGlnAlaGluGln...................... LysMetLeuTyrGlnValAlaGlyIly 313

1788 AAGGATAAATCCAG...........GCACATGGGCGCACTGTTGACCCCACCTTCA 1831
     |||||         ::::::|||
313 sLysMetIleGluAsnLeuGlnArgSerGlyLysLeuProProAsnLeu. 329

1832 CGAGTCGATGCAGGCACCACCTTCTGCTTGAGAGTCAACTCACCAGGCGTAT 1881
     |||||:
330 .......................... AspSerGlnInGlnLeuLeuTyr 337

1882 GCACAGGCCAATAAGACAAAGAAGAGATGTGTTACTTGCACACTGGTTTAAA 1931
     :::||:::
338 IleLys................................. LysTyrMetAs 343

1932 CACACTCTTAAAAAAAAAAAAAAAAATGTACCCGCCGACCGAGCCCTTCACT 1981
     ||||| :::  :::|||
343 nGlnMetValLeuLysLysMet............. GlnGlnLeuGlyI 355

1982 TCCCAGTCGCCTTCCCACCAGGAGAAAGGCCATGTTCACAGACATAATTATT 2031
     |||||       |||
355 leProIleProMetProGlySerAlaGlnGlnGlnSerGlnLeuLeuGln 371

2032 TCTGTGACTGGGATTTGTTGTAATGTGACAGAGAGAGATGACCTTAAAATTAAATGGC 2081
```

```
 372 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnArgGlnGlnGl 388
     ::: ::                             :::
2082 TTATTTGGCAGGTGCCAAAATATACGGGTTATCTATCCGCAGCAACACAG 2131
     ||||||| ||| |||             ||::||:||||||  ::
 388 n..IleProGlnVal.GlnGlnAlaProLeuGlnGlnProGlnValGlnPr 404

2132 TCCTCAATCTGTAAAGAGACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAG 2181
     ::||||||| ||:    :||||||| || ::|||||| |||
 404 oPro.............................GlnSerGlnGlnA 410

2182 TCGAGGATACCCTGTGTCAACCCCCAGTCCTTCCCGACATTCTTCTGGG 2231
     :::||       |||
 410 laGln......................................SerArgArg 414

2232 AAACTTTGAGGCACTGAGGCAGGATTCAGTATAGT.GCTACACGGCATTCA 2281
     :::  :::::                    |||
 415 GlnAlaMetGlyMetLysProThrProVal.............IlePr 426

2282 GTCTCGCAGGATCCCATTTGCCCCTACCCAGCCATTTACTTTTAAATCTTTTA 2331
     ::||||||| ||: ||| |||
 426 oAsnAlaGlyValMetLysProGlnProGln......................... 436

2332 GATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTAT 2381
     :::|||| |||          :::||||||||:::::
 437 ........GlnThr.ProGlnAsnIleGlnProMetMetGlnGlnGl 450

2382 AAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATGTCCACGCCTT 2431
     :::||||||:::||                 ||| :
 450 nSerSerProProToProGlnGlnGln......................GlnGlnG 461

2432 CTTCCCAAAGACCCAGAATTGAAGACGTACCACCTCCCACT......... 2472
     ::::||||||:::|||    |||||
 461 lnGlnGlnSerLysIle...SerValProArgProThrGluGlnAsp 476

2473 ..........AAAAAGCTAACTCACAGAATTGACC.......CCTTTTGT 2504
     ||||||||||::::  |||:::||            |||
 477 MetLeuAlaLeuLysGlyLeuAsnAlaGluValThrLysAsnProValGly 493

2505 GCTT.................TTCACTGGATTCGAGCCTGTCCAG......... 2532
     |||                 ::|||:||||
 493 sLeuAsnAspIleThrAsnArgLeuThrAsnAspGlnLysGlnIleIlea 510

2533 ....................AAGTGCACACCTCATTGCCAGCAAAGTGACTCGGCA 2629
     ................                ::|||:::||||||
 510 rgSerLysLeuHisAlaAsnGlnGlnLeuPheSerSerValGluSerPhe 526

2545 ATTAAGAAGCTCTACATTCTCTGCTGGAGGTTCGCGGAGCTGTCGCACAG. 2592
     |||                                   |||
 527 IleProThrLeuTyrMetLeuThrArgAsnGluHisIleArgGlnLeu 543

2593 ...............AAGTGCACACCTCATTGCCAGCAAAGTGACTCGGCA 2629
     :::||||| |||::                :::|||||:::||||||
 543 uLeuGlnIleArgMetLeuThrArgGlluIleMetGluLysAlaValArgG 560

2630 CCGTGAAGTTCCTGACGGCGATTTCTGTCGCTGAAGCACATAGTGACGCCA 2679
     || |||||||                     |||
 560 lyVal..PheLeuValGluProAsnValValAspLysValIle...... 573

2680 GAGTGGCCTGGAAGAAATGCTTCAGGTCTGCAAGTTCATTCGATGAGCAGAA 2729
     || |::||||                ||||||||||:::
 574 .........PheArgTyrGlnLysTyr................. 579

2730 CTACATTCTCCGAGATGCTGAGGCAGAAGTACTTTCTTCTTTCAGCTTGG 2779
     :::||||||                :::||||::::||
 580 ...................TyrGluTyrThrLysG 585

2780 AAGAATCCTTAAAAACGG 2796
     ||:::    |||:|||
 585 luGlnLeuLeuArgArg 590
```

OM of: US-09-664-641-10 to: PIR_71:*   out.format : pfs

Date: Jun 4, 2002 5:36 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool/US09664641/runat_04062002_110932_9184/app_query.fasta_1.3697
-DB=PIR_71: -QFMT=fastan -SUFFIX=rpr -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH-0.100 -LOOPCLI-0.000 -LOOPEXT=0.000 -QGAPOP-4.500
-QGAPEXT=0.050 -XGAPOP=0.000 -XGAPEXT-0.500 -FGAPOP-6.500
-FGAPEXT=7.000 -YGAPOP-10.000 -YGAPEXT-0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN-15 -MODE-LOCAL -OUTFMT=pfs -NORM=ext -HBAPSIZE=500
-MINLEN-0 -MAXLEN-200000000000 -USER-US09664641.@CGN1_1_173
-NCPU-6 -ICPU-3 -LONGLOG -DEV-TIMEOUT-120 -WARN_TIMEOUT=30
-NO_XLPXY -WAIT -THREADS-1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: PIR_71:*
Database sequences: 283138
Database length: 96089334
Search time (sec): 144.160000

score_list:
Sequence   Strd Orig   ZScore   EScore   len ! Documentation

Sequence	Strd	Orig	ZScore	EScore	len	Documentation
pir2:T24877	+	524.00	638.78	1.6e-28	1074	! hypothetical protein T13F2.3a
pir2:T24887	+	524.00	638.76	1.6e-28	1076	! hypothetical protein T13F2.3b
pir2:T37978	+	314.50	377.81	6.7e-14	878	! multiple BRCT domain containing
pir2:G85039	+	265.50	317.60	1.7e-10	765	! hypothetical protein AT4g03130
pir2:T49721	+	229.00	271.05	6.3e-08	831	! hypothetical protein B23L21.340
pir1:RGHYS5	+	223.50	263.37	1.5e-07	905	! regulatory protein SNF5 - yeast
pir2:S69205	+	220.50	257.16	2.6e-07	1180	! stripe a/b protein - fruit fly
pir2:T14577	+	219.50	253.96	3.2e-07	1457	! protein kinase YaKA (EC 2.7.1.
pir2:S37876	+	213.00	253.07	8.1e-07	738	! glutamine-rich protein YKL054c
pir2:T13167	+	210.50	246.05	1.3e-06	1010	! lola-like protein - fruit fly
pir2:S44898	+	204.50	238.61	3.3e-06	1000	! ZK1236.3 protein - Caenorhabdi
pir2:A38346	+	200.00	244.50	5.3e-06	230	! ultra-high-sulfur keratin 1 - m
pir2:H38346	+	198.00	244.27	7.3e-06	223	! ultra-high-sulfur keratin 2 - m
pir2:T14004	+	196.50	225.54	1.3e-05	1390	! trfA protein - slime mold (Dic
pir2:T03455	-	196.50	213.84	1.6e-05	4957	! ALR protein - human
pir2:T03454	+	196.50	213.29	1.6e-05	5262	! ALR protein - human
pir2:T13828	+	195.50	216.64	1.8e-05	3190	! CREB-binding protein homolog
pir2:V15651	+	192.50	238.94	1.7e-05	188	! DNA-binding protein C27A2.5 -
pir2:A34599	+	192.00	234.45	2.0e-05	286	! hypothetical protein C27A2.5 -
pir2:T24272	-	191.50	238.95	2.0e-05	164	! hypothetical protein T01B7.8 -
pir2:S60066	+	189.50	222.37	3.6e-05	756	! probable protein kinase YNL161w
pir2:T30608	+	188.50	217.07	4.5e-05	1173	! proteophosphoglycan homolog 6I
pir2:H53052	+	188.50	215.55	4.7e-05	1383	! guanine nucleotide exchange fa
pir2:B53225	+	188.00	219.16	4.6e-05	873	! ecdysone-induced protein E74A
pir2:S25365	+	188.00	218.23	4.7e-05	966	! CYC8 protein - yeast (Saccharom
pir2:T18235	+	186.50	214.78	6.2e-05	1145	! transcription activator GAL11
pir2:T11670	+	186.00	217.16	7.7e-05	2447	! hypothetical protein T13H2.3
pir2:S01992	-	185.50	224.28	5.9e-05	356	! glutenin low molecular weight c
pir2:T18975	+	184.50	230.86	5.9e-05	152	! hypothetical protein C06A1.6 -
pir2:T2431	-	184.50	205.94	0.0001	2129	! hypothetical protein Y79H2A.3
pir2:T05923	+	183.50	221.32	8.2e-05	374	! glutenin low molecular weight c
pir2:T03851	+	183.00	208.39	0.0001	1424	! thyroid hormone receptor activ
pir2:T00661	+	182.50	207.80	0.0001	1417	! hypothetical protein F3I6.24
pir2:I46592	+	182.00	220.13	0.0001	347	! involucrin - pig
pir1:JQ1150	+	181.50	213.75	0.0001	648	! protein kinase (EC 2.7.1.37) ca
pir2:T02440	+	181.50	205.84	0.0001	991	! hypothetical protein At2g41450
pir2:T23107	+	181.00	209.62	0.0001	947	! hypothetical protein H20J18.1
pir1:S18946	-	180.50	224.87	0.0001	169	! ultra high-sulfur keratin 1 -
pir2:T06982	+	180.50	217.93	0.0001	359	! glutenin low molecular weight c
pir2:S19033	+	180.50	208.43	0.0002	1008	! transcription activator GAL11

pir1:EDBE23	+	180.00	211.48	0.0002	676	! immediate-early protein IER2
pir2:A53185	+	180.00	211.05	0.0002	708	! G-box-binding factor - slime
pir2:T08588	+	179.00	207.29	0.0002	930	! hypothetical protein L23H3.3
pir2:A55038	+	178.00	205.11	0.0002	1028	! DNA-binding protein ovo - f
pir2:S16356	+	178.00	203.59	0.0002	1213	! ovo protein - fruit fly (Dr

seq_name: pir2:T24877

seq_documentation_block:
hypothetical protein T13F2.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24877
R:Swinburne, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19947
A:Accession: T24877
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1074 <WI>
A:Cross-references: EMBL:Z81122; PIDN:CAB03354.1; GSPDB:GN00022; CHSP:T13F2.3a
A:Experimental source: clone T13F2
A;Gene: CESP:T13F2.3a
C:Genetics:
A:Map position: 4
A:Introns: 58/3; 79/1; 112/1; 163/3; 268/1; 388/1; 547/3; 628/3; 669/2; 741/1;

alignment_scores:
         Quality: 524.00   Length: 692
         Ratio: 1.444   Gaps: 24
Percent Similarity: 52.457   Percent Identity: 25.000

alignment_block:
US-09-664-641-10 x T24877

Align seq 1/1   to: T24877   from: 1 to: 1074

1350 ACAGACTCTTTCAGCACC.........TACGAAAAATTTAGAACACACAGGTGA 1393
     ::: ||| |||:::: | ||| :::
347 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnIleGlyGlnPr 363

1394 ATCACAGCCACCAGCAGACACACAAAATGCCAATGCAGTGCGTGTTTAGCCAA 1443
     ::: ||| |||::|||::|||:|||::|||::|||
363 GlnThrAlaProProGlyMetArgGln... TyrIleSerProG 376

1444 GTGAAAGTGACTCCAGACACACATCCTACAGCA.........GCAGCA 1484
     :: ::||| ||| :::||| ||| ::: |||
376 LnAlaGlyGlnGlnGlnProThrProHisArgPheAlaPro 392

1485 GCAGCCCAGCACAGCAGCAGCAGCACCCCGGTTTTACACCTTCAGCCCC 1534
     |||||::: ||||:::|||::|||::||||
393 ProGlyProSerGlyThrThrSerSerProHisPheAlaProGlyAlaPr 409

1535 AGCAGATAAT.............GCAGCTCCAGCAGCAGCAGCAG 1566
     |::: |||:|||||||||||||
409 OGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnSer 425

1567 CAGCAG.............CCGCGCAT... 1608
     |||||| ||||||
426 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 442

1573 .............ATCTCTCAGCAACCTTACCCCCACCAGCAG 1599
     :: ||||||::|||||
442 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 459

1600 ....CCGCGCAT... 1608
     ||||||
459 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 475

1609 CCATTTTCACAGCAGCAGCAGCAGCAACACCAGCACC 1646
     ||| ::: ||||||::|||||
476 AlaGlyAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetLeuAs 492

US-09-664-641-10.rpr

A;Residues: 1-1076 <WIL>
A;Cross-references: EMBL:Z81122; PIDN:CAB54313.1; GSPDB:GN000022; CESP:T13F2.3b
A;Experimental source: clone T13F2
C;Genetics:
A;Gene: CESP:T13F2.3b
A;Map position: 4
A;Introns: 58/3; 79/1; 114/1; 165/3; 270/1; 390/1; 440/1; 549/3; 630/3; 671/2; 743/1; 90

alignment_block:
US-09-664-641-10 x T24887

Align seg 1/1 to: T24887 from: 1 to: 1076

alignment_scores:

Quality:	524.00	Length:	692
Ratio:	1.444	Gaps:	24
Percent Similarity:	52.457	Percent Identity:	25.000

```
1350 ACAGAGCTCTTTCAGCACC..........................TACGAAAAATTTAGAACAGCAGGTGA 1393
     : ||||                                      :::||  ||    ||:|  |||
 349 ThrProThrPheSerSerProAlaHisGlnMetPheArgGlnGlyGlnPr 365

1394 ATCACAGCCAGCAGGAGACAATACAAATGCCAATGCAGTGCTGTTAGCCAA 1443
     |||||||||||||||||   ||  ||    |:::  ||| |||
 365 oThrAlaProGlyMetArgGln.....................TyrIleSerProG 378

1444 GTGAAAGTGACTCCAGAGACACACACATGCTACAGCA...........GCAGCA 1484
     |:|||  |:|              |||  ||            ||
 378 lnAlaGlyGlnGlnProThrProGlyThrProHisArgPheAlaPro 394

1485 GCAGCCCAGCAGCAGCAGCAGCAGCAGCACCCGGTTTTACACCTTCAGCCCC 1534
     |||||||||||||||||||||||||||||        ::::||||
 395 ProGlyProSerGlyThrThrSerProHisPheAlaProGlyAlaPr 411

1535 AGCAGATAAT...........GCAGCTCCAGCAGCAGCAGCAG 1566
     | :||||                |||  ||||||||||||
 411 oGlyAlaAspLeuArgSerProSerLeuMetSerPro.AsnGlnGlnSer 427

1567 CAGCAG..................................................... 1572
     :||||
 428 GlnGlnHisThrProValLeuProArgSerLeuHisAsnGluMetThrSe 444

1573 ...................ATCTCTCAGCACCTTACCCCAGCAG. 1599
                       :::  ||||||| |||||||||||
 444 rHisGlyMetArgProIleAlaLeuProSerGlnProLeuProGlnThrG 461

1600 ..........................CCGCGCGAT... 1608
                             ||||||:::
 461 lnGlyThrProGlnProIleLeuSerAlaSerSerLeuProProGlnSer 477

1609 ...........CCAATTTCACAGCACCAGCAGCAGCAGCAACAGCAGCCACC 1646
            :::  ||||||||||||||||||||
 478 AlaGlyAlaAlaThrGlnPheValTyrArgProGlnGlnProAsnGlnMetAs 494

1647 ACCATCCGCCTCAGCAGCATCACGCTTTTTGGACATGAGTCCA........ 1686
     ::||||||||||||||||||||          ||||||
 494 nThrMetAsnGlnMetAsnGlnMetHis.....HisAspValAsnProPheP 510

1687 ....GCAGTGGAGAGTTCCACAGAGAAGGCTTCTTATTTGGAATGTGTGTTT 1731
         ::||||||||  |||||||||||||||| |||  ||
 510 rolleGlnMetLeuAsnValAspProGluLeuTyrLeuThrGlyIleHisPhe 526

1732 GCAATTCGCGGATTATTCCAGAGACAGATGTCATAAGCAACTGCTGGCCCAC 1781
     |:::|||||||||||||||||||||||| |||
 527 HisCysPheAspSerGlyLeuPheGluAspLeuAspArgLysAs 543

1782 CTGGAAAAGGATAATTCAGGACACATGCGGCACTGTTGAC..........C 1822
     :::||||||||||||||||  |||  ||||||||||
 543 nLeuGluPheMetIleLysTyrHisGlyAspIleGluPheGluAlaS 560
```

seq_name: pir2:T37978

seq_documentation_block:
multiple BRCT domain containing protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T37978; T40564
R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z21759
A:Accession: T37978
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-878 <MUR>
A:Cross-references: EMBL:Z69909; PIDN:CAA93789.1; GSPDB:GN00066; SPDB:SPAC19G10.07
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21937
A:Accession: T40564
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-878 <WOO>
A:Cross-references: EMBL:AL096788; PIDN:CAB66668.1; GSPDB:GN00067; SPDB:SPBC582.05c
A:Experimental source: strain 972h-; cosmid c582
A:Genetics: <GEN1>
A:Gene: SPAC19G10.07
A:Map position: 1
A:Introns: 37/3; 83/3
A:Map position: 1
A:Genetics: <GEN2>
A:Gene: SPBC582.05c
A:Map position: 2

---

alignment_scores:
    Quality: 314.50   Length: 1194
    Ratio: 0.758   Gaps: 35
    Percent Similarity: 34.757   Percent Identity: 17.169

alignment_block:
us-09-664-641-10:
US-09-664-641-10 x T37978

Align seq 1/1 to: T37978 from: 1 to: 878

```
2620 GTGACTCGCACCGTGAAGTCTTCCTGACGCGGCGGATTTCTGTCGTGAAGCACAT 2669
     ||:||||:    |||     |||||||:|||
 872 GlyAspArgSerLeuValLeuLeuGluSerIleIleArgGlyLysAsnIl 888

2670 AGTGACCCAGAGTGGCCTGGAAGAAATGCTTCAGTGTGTCAGAAGTTCATTG 2719
     |||||||||::   |:|||||:    |:::||
 888 eMetAspProGluTrpIleValAspSerTyrLysGlnLysMetTrpLeuA 905

2720 ATGAGCAGAGAACTACATTCTCCGAGATGCTGAGCAGCAGAAGTACTTTCTCT 2769
     :::||:||||   ||||:|::|||    |||
 905 spThrLeuAspThrPheLeuHisAspLysGlyLeuGluLysGluPheAla 921

2770 TTCAGCTTGGAAGAATCCTTAAAACGGGCACACGTTTCTCCACTCTTTAA 2819
     :|:::||:    |||||:    :::||
 922 TyrAsnCysLysArgSerValLeuArgAlaArgAsnLysProValPheGl 938

2820 GGCAAAATATTTTTACATCACGACTGGAAATCTGCCCAAGTCTCTTTCCACTA 2869
     ||||:|||||    ::::||
 938 uAspIleGluPheHisValThrArgPheValGluProAsnGlnAsnAspL 955

2870 TGAAGGCAATCGTAGAGTGTGGCAGGAGGAAAAGGTCTGTTATCCAAGCAGCCA 2919
     :::||:|||    |||:::|||
 955 euValArgLeuIleGluLeuGlyGlyGlyAsnValHisSerGlyLysPro 971

2920 TCT.........TTCCGGAAGCTCAGTGGGACACAAGCAGAACTCGAGTTT 2960
     :::   ||:|||| ||| |||
 972 AspProLysGlyTyrLeuAlaLysCysValGluThrArgGluGln 984

2961 GTCCGAAATAATATTTAATATCCTGTGAAAATGACCTTCATTTATGCCGAG 3010
     :::|||||:|||||||:|||:::|||::|:|||
 985 ...ProPheIleIleSerCysGluAsnAspLeuArgPheLeuSerT 1000

3011 AATATTTTCCAGAGGCGATAGATGTTCACAATGCAGAGTTCGTTCTGACT 3060
     :::::|:||::|||||:|||||
1000 yrLeuAlaGlySerLysLeuProIleTyrAsnValAspLeuValLeuPhe 1016

3061 CGACTGCTCACTCAAACGCTGAC 3084
     :::||:||| :::::|||
1017 AlaMetLeuArgGlnIleIleGlu 1024
```

```
144 TACGGATCTCTCGGCCCGCCCCGGCCCCCTTCAAATCACCTCTAGGCACAGGTTAT 193
    :||:||||    :::|||
 25 TyrSerAlaLeuGluGluPro...AsnGlnAsnGluGluLeuArgGlyLysGluGlu 40

194 TCAGCTTCTCAAGGCGTGAAAAAGCGAAGAAGTTCCTTACAAT...GCAC 240
    :::|||||||||||:||:::
 40 uPheIleLysAsnAspGlyLysAlaLeuGluSerPheProTyrAspIleLysL 57

241 TAGCCTCACACATAATCTCAGAGATGGGACAAATCCAGAGAGGTGGGACAGA 290
    :::|||:::|||||    ::|||||:::|||
 57 euAlaThrHisValIleCysAspAspPheSerSerProAsnValGlnGlu 73

291 GCTCGGGAAGTCTTGACTTACCTGTTGTAAAACCTTCTTGGGTGATTCT 340
    |||:::|||||    :::|||||:::|||:::||||
 74 GlySerLys...ArgSerLeuArgLeuAlaLysThrAsnTrpIleArgAs 89

341 GTCCGGTCAGTGTGGAACGCTTCTGCCAGTAAAATGGTTTTCTCCAGAAT 390
    ||:::|||||||||||||:::
 89 pCysValAspLysAsnThrLeuLeuAsnTyrSerPheTyrSerCysAsnp 106

391 CANGTCAGATTTTTTTGGAATCACACTGCCTGCCTTCTCAGGGTGTTGAT 440
    :::|||   |:||::|||||||
106 roTyrLeuLeuPheLysGlyIleCysAspAspPheSerSerCysGln...IleAsp 121

441 ACAAGCTGGAGCTTCTTTGTTCTTGGAGTCTTCCAGAGCTGCTCCCAACGACGAGG 490
    :::||:::|||||||||:::
122 SerTyrGlnSerSerLeuIleAsp... 129

491 TAGGGAAGGGAGCGTTGTCCACCAGAAGTTGGAAGCACAGAGATCATCTG 540
    :||||||||
135 ..............................................P 135

541 CCTTCTTCTGACCCGGTGATTGATCCAGGCTGAGGCCTCTCGTTGTGTAATGTG 590
    ||||||| :::
130 ..........AspAla.LeuGluThr.............. 134

591 CTGGGTGTCATCTGAAGACAAGAGTGCCCTGTGGGCTTTGGTTACGTTCT 640
    |||||||:::
135 ...........................................P 135

641 ATGGGGGAGAATTGCCAGCTAACCCTCAATAAGAAAATGCACGCATTTGATT 690
    :||||||||
135 ..............................................P 135

691 GTTCCAGAGCCAAAAGGGGGGAAATACGAAGTGCTTAAAGCGA..... 735
    ||||||| :::
152 ThrThrTyrSerGlyMetGlyAlaLysCysLysLysValLeuAspLysProSe 168

736 .GCAAGTATTAAAATTGTGACTCCTGACTGGGTTCTCGGATTGCGTATCAG 784
    |||||||||
168 rLeuSerIleLysLeuIleHisProGlnTrpLeuLeuAspCysLeuGlnP 185

785 AGAAAACCAAAAAAGGACGAAGCATTTATCATCCTCGTCTGATTATTTAT 834
    :||||||||
185 heGlyGlnLeuLeuAspGln.............. 191

835 GAAGAGGAAGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAAAATGAGGA 884
    :::
191 ............................ 191

885 ACAAGATTCTCAGAATGAATGAGGGTAGTACAGATGAGAACTCAAGCCCTGCCA 934
    ............... 191
```

```
 191 ..................................................................... 191
 935 GCTCTCAAGAAGGGTCTCCTCCTTCAGGTGACCAGCAGTGTTTTCACCTAAATCC         984
                |||
 192 ...........................................AspProTyrLeuPheProAsnPro     199

 985 AACACTGAAAAAATCTAAAGGGCAGAATTAAATCTTTGATCATTCTTCAGATTC        1034
         ::://                   :://|||
 200 SerTyrLysLys...........................AsnAspSe       206

1035 ATCACCGGCAAAAACAGGAGCAGAAATTTAAACTGGACCCGGCCGAAGTCC       1084
     |||  ::://|||         |||
 206 rSerIleSerLysAlaGlu..................................       212

1085 CACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGGAGGCCTGGG       1134
     ...................................................       212

1135 TTTGATTAACTTGTGTCTGTCCCAATGTCCCACCGGTTCCAGGGTAACATTTTGCC      1184
             ::://|||::://|||                   ::://|||
 213 ...........................................ProThrSerLeu...       216

1185 CCCTGAGGTCCGGGGTAATTTAAICGGCTGGCTGCTCGCACAAAACCTCCAAAGTT       1234
     ...................................................       216

1235 CTGAAAGATCAGAAAATGATAGCTACCTGGAGTCCAGCTGTACGGACACTG       1284
     ...................................................       216

1285 AGGAATATTTACTTAATAATAATGCCTGACATTT........CACCAGATGAACCG      1325
     |||||||||::://|||:://|||
 218 ArgAsnValLeuHisGlyLysArgIleTyrPheSerAsnAspLeuAsnLe       234

1326 GCCATCAAATGTGAGCACATATCTTACAGACATCTTTCAGCCACCTACGAAAA        1375
     :://||||| |||  ||||||
 234 uProThrAsnPheArgHisSerLeuGlnLys...................       244

1376 ATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACTACAAAATGCCAAT       1425
     ...................................................       244

1426 CCACTGCTCTGTTTAGCCAAGTGACTCCAGAGCACACACATGCTACA        1475
     ...................................................       244

1476 GCACGACGAGCAGGCCCCACCAGCAGCAGCAGCAGACCCCGGTTTTACACC       1525
     ...................................................       244

1526 TTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGATC        1575
     ...................................................       244

1576 TCTCAGCAACCTTACCCCCAGCAGCCGCATCCATTTTCACAGCAGCA       1625
     |||                         ::://|||
 244 ...........................PheSerValGlyIle         249

1626 GCAGCAGCAACAGCAGCCACCACCATCGGCCTCAGCAGCATCAGCTTTTTG       1675
     |||||
 250 GlyAlaLysIleAla..........GluSerIleAsnAspCysAspIlePh     263

1676 GACATGATTCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGGGATGT       1725
     |||||
 245 ...........................FheServalGlyIle        249

1726 GTGTTTGCAATTCGGGATTATCCAGCAGATGTCTGATAAGCAACTGCT       1775
     ||||||
 250 GlyAlaLysIleAla.........GluSerIleAsnAspCysAspIlePh     263

1776 GGCCACCTGGAAAGGATAATTCAGGCACATGGCGGCCACTGTTGACCCCA       1825
     : ||||||
 528 oProSerProLeuLeuValGlyThrProSerLysGluSerLeuLysGluA     545
```

```
 263 elleGlyLeuLysArg.............................AspThrIleGlu...    272
1826 CCTTCACGAGTCGATGCAGCGACCACCTTCTCTGTGAGAGTCAGCAGG     1875
                                                  ::::::
 272 .............................................................    272

1876 GCCTATCCACAGGCAATAAGAGAAAGAGAAAGAGATCGTTCACACTG       1925
     :::                 |||     |||
 273 ...PheAsnLeuAlaSerAsnLysAsnThrThrIleGlyThrIleSerIr    288

1926 G....TTAAACACNGTCTTAAAAAAAAAAAAAAAAAAATGGTACCGCGCACCGAG    1972
     |||||
 288 pLeuLeuAsnLeuPheValLeuGlySerIrpLysSerProLeuGLeuAsnA    305

1973 CCCTTCACTTCCCAGTGGCCTTCCACCAGGAGGAAAGCCATGTTCACAG     2022
     |||||||||::://|||           |||||         :::::
 305 laLeuHisTyrPro........PheProSerValGlyPhe...LeuLysAsp    318

2023 CATATTATTTCTGTGACTGGATTGTTGATAGTGACACAGAGATGACCTAA    2072
     :::||||||::://|||:://|||               ::://||:::
 319 GlnMetValAlaValThrAsnTyrThrAspAlaAlaArgIleTyrLeuGl    335

2073 ATTAATGGCTTATTTGGCAGGTGCCAAAATATAGGGTTTATCTGATCGCGCA    2122
     :::                    |||||| |||||:::::  :::
 335 uLysLeuLeuLeuLeuAlaCysGlyAlaThrTyrThrLysAspLeuLysSerProT   352

2123 GCAACACAGTGCTCATCGTGTAAAGAAACAACTGGTTTAAAGTAATAGAAAA   2172
     ::|||||||::://|||:://|||              |||   |||||
 352 hrAsnThrLeuLeuIleAlaAlaSerSerTyrGlyLnLysTyrGlyVala   368

2173 GCCAAAGACTGGAGGATACCCTGTGTCAACGCCCAGTGGCTGGCGACAT   2222
     ::||||||||::://|||:://|||:::
 369 AlaLysValTrpAsnIleProThrValHisHisSerTrpLeu......   382

2223 TCTTCTGGGAAACTTGAGGCAGACTGAGCGAGATTCAGTAAGTCGGCTACA   2272
     :::::::::|||::://|||:::::||||||:::::  :::
 383 ...........TyrSerSerPheLysAsnLeuSerSerGlnAlaPhe   395

2273 CGGCATTC............................................   2280
     |||
 395 .............TyrSerSerPheLysAsnLeuSerSerGlnAlaPhe   395

2281 hrAspPheProValProLeuAspApSerTyrMetAspPheHilePhePro   411
     ...........................AGTCTGCAGGATCCA.           2295

2295 ...................................................   2295

2296 TTTGCCCCTACC...................CAGCATTTAGTTTTTA       2322
     |||
 445 erValSerSerIleLysGlyAsnLysThrAsnGluGluLeuGluLysGlu   461

2322 ...................................................   2322

2323 ....................................AATCTTTTA      2331
 478 rPheSerAsnGlnSerAlaAspLysGlySerSerLeuAlaLaGlvAspA    495

2332 GATGCTTGGAGAGTT..................................   2346
 495 spArgAsnAspGluGlySerThrIleThrGlyValAsnArgGluLeuGln   511

2347 .......CCCTTAAAAGTGTCTGCAGAGTTGTTGATGAGTATAAGA.....   2385
     |||||||::://|||::://|||
 512 AspGluGlyArgLeulleAspAlaLysSerSerLysThrAsnThrPr      528
```

seq_name: pir2:G85039

seq_documentation_block:
hypothetical protein AT4g03130 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C;Accession: G85039
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488
A;Accession: G85039
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-765 <STO>
A;Cross-references: GB:NC_001268; NID:g9270183; PIDN:CAB77798.1; GSPDB:GN00140
A;Gene: AT4g03130
A;Map position: 4

alignment_scores:
Quality: 265.50   Length: 910
Ratio: 0.676   Gaps: 34
Percent Similarity: 43.187   Percent Identity: 19.890

alignment_block:
US-09-664-641-10 x G85039

Align seq 1/1   to: G85039   from: 1   to: 765

```
 766 GTTCTGGATTGC..........GTATCAGAAGAAAACCAAAAAAGGACGAAGGCATT  809
     ::::|||  |||         :::::|:::  |||||||||:::
  26 IleLeuGlyCysGlyGluIleGluAspThrGluAlaIleAlaAspGluAlaIle  42

 810 TTAT..........................CATCCTTCGTCTGATTATTATGAAG  838
     :                             |||           |||:|:|:
  42 uValLeuAsnSerSerLeuLeuGluThrProPheGlnValLeuTyrAspA  59

 839 AGGAGGAA.......................GAGGAGAAGGAAGAAGGAG  864
     :::::::                       |||||||||:::
  59 spThrGluLeuValAspAsnGlyAsnCysGlyAspGluGluSerLeuAsn  75

 865 GAGGAAGTAGAAAAATGAGGAACAA...............GATTCTCAGAATGA  902
     :::||| :::  :||:||          |||||||||:::
  76 GlnGluCysTyrGlyLeuAspGlnValAlaValAspAspAspSerAspSerAspPro  92

 903 GGGTAGTACAGATGAGAGTCAAGCCCTGCCAGCTCTCAAGAGAGGGTCTC  952
     :::::                      |||           |||  ||||
  92 uAspGlySerValAlaAlaGlyValuProMetCysLeuuProGluLeuSerP  109

 953 CT...TCAGGTGAC............................  963
     ::  :::||||||                    ||||
 109 roLysAlaGlyLysAspIleLeuLeuGluSerArgAspHisGlyGlu  125

 964 ...CAGCAGTTTTCACCTAAAATCCAC............  987
     :::  :::|:|||||                   ||
 126 CysGlnThrGlyLysGlyGlnLeuSerAsnCysAspThrValThrGlyPheGl  142

 988 ..................ACTGAAAAATCTAAAGGGGAA...TTAATGTTTG  1018
     :::                |||      |||  |||   ||| |||
 142 nGlySerLeuuArgIleThrAlaSerSerHisSerGlyProGlyLeuAspPheL  159

1019 AGATTCTTCAGATTCATCACCGGAAAACAG.................  1050
     |||||:::::  :::    ||||
```

```
159  euAspSerGlnGluProGlyGluSerThrGlnAlaAspAlaLeuGlyPhe 175
1051                          GAGAGAAATTTAAACTGGACCCCGCGAAGT
                              ::|||||||||||||:::|||||:::||||
176  ValAspHisPheLeuMetAspLeuSerProValAsnLeuSerProValAsple 192
1083 CCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGAGAAGAGGCCTG 1132
     :||:          ||||||||||||||||||||||||||||||||
192  uProValAsnSerCysArgArgLysSerProValSerGlyAlaLysG 209
1133 GCTTGATTAACTGTGTGCCCAATGTCCCACCCGTTCCAGTTAACATTTG 1182
     :|||          ||:          :::|||
209  lyCysGlnSerLeuAlaLysArgIleMetSer 219
1183 CCCCCTGAGGTTCCGGGGTTAATTTAATCCGCTGGACAAAACCTCCAAAG 1232
     :::          ||:|||
220                             ArgSe 221
1233 TTCTGAAAGATCAGAAAATGAATAGCTACCTGG 1263
     :::                        |||
221  rProThrArgLysMetSerValPheAspTrpSerSerAspGlnCysAspV 238
1264              AGTCCAGCTGTACCG       ACACTGAGG 1287
                  ||||||               |||
238  alSerAspProLysThrSerProValThrArgGlnAlaSerIleThrCysSer 254
1288 AATATTACTAATAATGCT             GACATTCAGCAGATGAACCG 1325
     ::|||                          ||:|||
255  LysIleArgGlnValAlaGlnLysAspAlaProPheGlyValValLysSAr 271
1326 GCCACATCAAATGTAGTACCACATATCTTACACGACTCTTTCAGCACTTACGAAAA 1375
     :::                            |||||:::|||:::
271  gSerArgAspProCysAspIleArgGlyValSerSerHisProThrGlnA 288
1376 ATTTTAGAACACGAGGTCGAATCACAGCCAGCAGCACATACAAAATGCCAAT 1425
     :::|||                        |||||||||:::
288  rgValMetLys                    AsnSerSerAlaLys 296
1426 CCAGGTGCTGTTAGCCAAAGTGAAAGTGACTCCAGACACACAATGCCTACA 1475
     :::                            |||||||||
297  HisLeuLysMetGluGlnAlaSerGlyLeuSerGlyIleMetPheIl 313
1476 GCAGCAGCAGCAGCGCCAGCAGCAACCAGCAGCACACCCGGTTTTACACC 1525
     :::                             |||
313  cSerGlnLysAspValAlaGlnIleuGluAspLysLysA           ProLysGluProL 329
1526 TTCACCCCCAGCAGATAATC                          CAGCTCCAG 1554
     :::                                           |||:||
329  euGluProGluGluuAspPheIleAspIleGlyIleAsnThrGlnIleAla 345
1555 CAGCAGCAGCAGCCAGATCTCTCAGCAACCTTACCCC 1593
     :::|||                        |||
346  AlaGluuAlaMetSerAlaLeuuValThrArgAlaProCysThrThrGluGluAl 362
1594        CAGCAGCCGCGCCATCCATTTTCACAGCAGCAGCAGCAACACAGC 1639
            |||          |||          |||||||
362  acCysGlnSerAspProIleProGlySerValSerGlyMetArgAspGlnv 379
1640 AGCCACCACCATCGCCTCAGCACCATCAGCTTTTTGGGACATGATCCAGCA 1689
     :::                           |||||||
379  alSerAsnLeuuSerArgAyrAsnAsnAspThrIleGluGlyValPro 394
1690 GTGGGAGATTCCAGAAGAAGAGGCTTCTAATTGGCAGTGTGTTTGCAATTGC 1739
     :::                           |||          ::||||:::
395             GluArgAspLysLeuuAsnGlyLeuPheSerAlaAl 406
1740 GGATTATCCAGAGCAGAGTCTCTGATAAGCAACTCCTGCCCACTGGAAAA 1789
     :::                          |||          |||||:::
406  aProHisArgGluArgAsnSerLysLysArgLysPheThrArgAsnGluG 423
1790 GGATAAATCCAGGCACACATGGCGGCCACTGTTGACCCCACTTCACGACTCGA 1839
     :::                         |||          ::|||
423  luArgThrArgAlaAsn                         ValSerValLysThr 433
```

```
1840 TGCAGGCACCTTCTCTGTGGAGAGTCAAGTCACGCGGTGATGCACAGGC 1889
     |||                       :::        ::||
434  CysLeuLeuAsnLeuuCysGluCysGluThrpArg      HisProArgA 446
1890 AATAAGAGAAAAGAAAAAGAGATGTGTTACTGCGACACTGGTTAAACACAGTCT 1939
     :::||||:::                |||||||||||          ::||
446  aLysArgSerArg            L 451
1940 TAAAAAAAAAAAAAATGGTACCGGCGGCAGGAGCCCTTCACTTCCCAGTTG 1989
     ::|||||||||:::|||         |||||||||::||:
451  euMetGlnArgHisHisHisValProProArgProArgSerTrp 463
1990 GCCTTCCACCCAGGAGGAGAAAGCCATGTCACAGCGATAATTATTTCTGTGAC 2039
     :::                        ||||||||||::||:::
463                             Val 463
2040 TGGATTTGTTGATAGTGACAGAGATGACCTAAAAATTAATGGCTTATTGG 2089
     :::                          |||          ::||:::
463                             463
2090 CAGGTGCCAAAATATACGGGTTTATCTATGCCCGACAGCAACAGCTCCTCATC 2139
     |||||||||                     :::|||||:::|||
464  .GlyAlaSerLeuuValLysAspArgSerGluThrAsnThr 476
2140 TGTAAAGAACCAACTGGTTTAAAGTATGAAAAAGCCAAAGAGTGGAGGAT 2189
     :::                          |||          ::|||
476                             476
2190 ACCCTGTGTCAACGCCCAGTGGCGTTGGCGACATCTCTTGGTGGAAACTTTG 2239
     :::                          |||||||:::|||:::
477                             LeuSerSerLeuArgGlySerLeuS 486
2240 AGGCACTGCAGGCGAGATTCAGTATAAGTCGCGTACACGGCAATTCAGCTGCAG 2289
     :::                          |||||||::::::
486  erGlyThrAr.GlnAla                       491
2290 GATCCCATTTGCCCCTACCCCAGCATTTAGTTTTAAATCTTTTAGATGCTTG 2339
     :::                          |||||||:::||:::
492             .SerSerCysGlnSerGlyValIleAspLeuuAsnValAlaTy 505
2340 GAGAGTT...CCCTTAAAAGTG                         2358
     :::      |||
505  rHisAlaSerProArgLysIleTyrAspGlySerHisGluSerProCysA 522
2359                        TCTGCAGAGTTGTTGATGAGTATA 2382
                           |||||||:::|||:::
522  snLysAspPheProArgLeuuPheLeuuGlnLysGluLeuuThrThrSerLeu 538
2383 AGACTACCTCCCAAACTG          AAACA 2405
     :::                         |||:::
539  GlyGlyProGlyLysIleGlyAspPheValTrpLysAspLeuuArgLysSAr 555
2406 GAATGAAGTAGCTAATGTCCAGCCCTTCTCCCAAAAGAGCCAGAATTGAAG 2455
     :::                        |||||||:::|||:::
555  gArgAsnGluuAlaHisValArg                         562
2456 ACGTACCACCTCCCACTCAAAAAAGCTAACTCCAGAATTGACCCCTTTGTG 2505
     :::                        |||          :::
563                             Val 563
2506 CTTTTCACT...GGATTCGACCCTGTCCAGGTTCAACAGTATAATTAAGAA 2552
     ||||||||||                    :::|||
564  LeuPheSerGlnAsnLeuuAspProGluThrValLysGlnGlnlyLysIll 580
2553 GCTCTACACATTCTTGGTGGAGAGTTGCCGGAGTCTGCACAGACAAGTGCACAC 2602
     :::                        |||          :::|||
580  eMetValArgLeuuGlyIleSerProAlaSerSerAlaAspSerPheArgH 597
2603 ACCTCATTGCCCAGCAAAAGTGACTCCGCACCGTGAAGTTCCTGACGGCGATT 2652
     :::                        |||          ||:|||
597  IsPheIleAlaAspArgPheAlaArgThrArgAsnMetLeuuGlnAlaIle 613
```

alignment_scores:
    Quality: 229.00   Length: 725
    Ratio: 0.727   Gaps: 21
Percent Similarity: 43.448   Percent Identity: 18.207

Align seq 1/1 to: T49721 from: 1 to: 831

alignment_block:
US-09-664-641-10 x T49721

seq_name: pir2:T49721

seq_documentation_block:
hypothetical protein B23L21.340 [imported] - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C;Accession: T49721
R;Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Hartmann, B.; Holland, R. Nyakatura,
submitted to the Protein Sequence Database, May 2000
A;Reference number: 225022
A;Accession: T49721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-831 <SCH>
A;Cross-references: EMBL:AL356172; GSPDB:GN00116; NCSP:B23L21.340
A;Experimental source: BAC clone H23L21; strain OR74A
C;Genetics:
A;Gene: NCSP:B23L21.340
A;Map position: 6
A;Introns: 31/3; 629/3

1156 GTCCCACACCGTCCCAGGTAACATTTTGCCCCTGAGCTCCGGCGGTAATTT 1205

757 LeuSerThrThrAspPrp....TyrLys 764
    :::||| |||:||
3070 ACTCAAACCCTGGACTATGAATCATATAAG 3099

742 ysGlyThrLysGlyIleHis......LysPheTh:AlaAlaCysLeuLeu 756
    :::||:|||:|| ::: :||:||
3031 .........GATGTCACATGCCAGTAG:C:CTCTGACTGGAGTGCTC 3069

725 eCysLeuProPheValAsnGlnGlyTyrThrValSerSerGlnPheGlnArgL 742
    ::|:| :||
3003 ATGCGAGAAATAATTTTGCCAGAGGCATA............... 3030

709 AsnPheProGluAspValLeuLeuSerCysLysGluAspArgAspPh 725
    ||||| :: ::::||||:||||||||:|||:||:
2953 TCGAGTTTGTCGGAAATAAATTTTAATATCCTGAAAATGACCTTCATTT 3002

696 alValAlaIleAla...........GluIleIleAlaAlaGluAspArg 708
    :::||| :::||| :||:||
2903 GCCAGCAGGGACAATACAAAATGCCAATGCAGTGCCTGTGTTAGCCAAGTGAAA 2952

679 sProSerArgGlyMetIleThrAspLeuValLeuSerCysLysGlyAspArgAspPh 696
    ||||||||||||||||||||:|||||||||||:||:||:||:
2853 CCCAAGTCCTTTCCACTATGAAGGCAAATCGTAGAGTGTGCCGAGGAGGAAAGG 2902

663 GlnIleIleProLeuLeuPheLysGlyPheLysValCysIleThrProSerIleLy 679
    |||||:|||||| ||||||:|||:||
2803 GTTTCTCCACTCTTTAAGGCAAAATAATTTTTTACATCACACCTGGAATCTG 2852

647 ysGlu...LysAspGlyPheCysLeuLeuThrSerLeuAlaArgAlaLys 662
    ||| |||||||||||||||||||||||| ||||||
2753 CAGAGTACTATTCTTCTGTTTCAGTCTTGGAGAAATCCTTAAAACGGGACAC 2802

630 nThrArgCysLeuLeuAspGlyLysSerTyrIleLeuArgAspIleLysL 647
    :::|||||||||||||||||||||||||||||||||| :::
2703 GTGTCAGAAGTTCATTGATGAGACAGAACTACATTCTCCGAGATGCTGAGG 2752

614 AlaLeuLeuGlyLysPheValValThrProIleTrpLeuGluSerCysAlaGl 630
    ||| :::|| |||||||||| :||
2653 TCTGTCCTGAAGCACATAGTGACCCGCAGAGTGGCTGGAAGAAATGCTTCAG 2702

---

71 IleProValAlaAsnSerAsnTrpIleLysIleThrLeuAlaArgAsnLy 87
    :::||| ::: :::
1206 AATGGCTGCTGGACAAAAACCTCCAAAGTTCTGAAAGATCACAAAATGATAG 1255

87 sValAla.......................................GlnValA 92
    ::||| :::
1256 CTACCTGGAGTCCAGCGTGTACGGACACTGAGGAAT........ATTACTAAT 1299

92 rgProTyrSerProAspProArgMetIlePheSerAsnValIleLeuThr 108
    ||||||| ||| :::
1300 AATGCTGACATTCAGCAGATGAACCGCCATCAAATGTAGCACATATGTT 1349

109 CysAlaAspIleProGluSerAspLysGluThrIleIleGlyAlaThrMe 125
    |||||||| ||||||:||
1350 ACAGACTCTTTCAGCCACCTAGAAAAAATTAGAACAGCAGGTGAATCACA 1399

125 tAlaLeuGlyGlyMetGluSerLysAspLeuHisArgGlnThrThrHisI 142
    :::||||:: ||:|||::||||||:||
1400 GCCAGCAGGGACAATACAAAATGCCAATGCAGTGCCTGTGTTAGCCAAGTGAAA 1449

142 le..Pr 142
    :::
188 ....Pr 188

143 ....CysAlaLeuSerMetAspHisGluLysCysGlnGluAl 155
    ::||||||||:||||:
1450 GTGACTCCAGACACACATGCTACGAGCAGCAGCAGCAGCCACCA 1499

155 aGlnLysLysAsnProLysCysLysIleValLeuProHisTrpPheAspA 172
    ||| |||:::||||||:::||||||:::
1500 GCAGAGCAGCACCCGGTTTTTACCTTCAGCCCAGCAGATAATGCAGC 1549

172 spCysPheArgLeuGlyArgArgIleSerGluAlaProTyrMetLeu... 187
    :::||||||::: :||||||| ||||||
1550 TCCAGGAGCAGCAGCAGCAGATCTCACCACCTTACCCCACCAG 1599

187 ....CysAlaAlaCysLeuLeu 187
    || ||
1600 CCGCCCCATCCATTTCACACAGCAGCAGCAACAGCAGCCACCACC 1649

188 le.........................Pr 188
    ::||
188 ...Pr 188

204 roProSerGluAlaValGluGlyAlaThrSerValIleProSerAlaAla 220
    ||| :::|||:::|||||| |||
1700 CAGAAGAAGGCTTCTTATTGGGATGTGTGTCTTTGCAATT......GCGGAT 1743

221 MetProGlyArgGluArgGluGlyIleValAlaPheAlaGlnLysLysysVa 237
    ::: :||||
1744 TATCCAGACAGCAGATGTCTGATAA:CAACTGCTGGCCACCTGGAAAGGA: 1793

237 lMetIleSerGlnAspLeuProIleAsnAlaArgLeuArgAsnIleLeuUa 254
    ||||||| ||| :::
1794 AATC.. 1797

254 snAsnLeuIleGluGlnSerAspGlyGluValValAlaValAsnAsp.. 268
    |||||| :::||||||
1798 .........CAGGCAATGCGCGCACTGTTGACCCCACACCTTCACG 1833

269 AspAlaCysSerAspMetThryrValCysGlnTyrArgAspGlyAspSpGluTyrrIl 285
    ::||| :::|||:||:||:||:||
1834 AGTGCGATGCACGCACCTTCTCTGTGAGAGTCAAGTCACGCGGGCGTATCG 1883

285 eArgAlaAlaGlnInGlyLyLysAspValGlyAsnLeuAlaTrpLeuTyrT 302
    ||||||| :::||||||
1884 ACAGGCAATAAGAGAAAAGAGAGATGTGTTACTGCACACTGGTTAAACA 1933

302 yrLeuIleValHisAsnGluTrpThrArgProThrArgArgLeuLeuHis 318
    ::||| ||| ||||||
1934 CAGTCTTAAAAAAAAAAAAAAAAAATGGTA...CCGCGCAGCACCCTTCAC 1980

318 ....Pr 318
    ::::|||:::::
1981 TTCCCAGTGGG:TTCCACCAGAGAGAAAACCCATGTTCACAGCACAATT.. 2028

```
319 'TyrProIle.........ProArgAspGlyIleProGlyPheThrGlyMetLy 333
2029 ATTTCTGTGACTGGATTGTTGATAGTGAGACAGAGATGACCTAAAATTAA 2077
         ::: :::|||||||         |||||  :
333 sIleThrLeuSerAsnTyrGlyGlyValTyrAspAlaArgValTyrLeuGluAsnL 350
2078 TGGCTTATTTGGCAGGTGCCAAATATACCGGTTATCTATCCCGCAGCAAC 2127
     |||||||  ||||||         |||  :::|||
350 eutIeThrAlaAlaGlyValAlaThrTyrThrLysThrMetLysAlaAspAsn 366
2128 ACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAGTNTGAAAAAGCCAA 2177
     |||||||||:::|||       |||||  ||||
367 ThrHisLeuIleThrAlaAArgMetAsnGlyValTyrGluLysCysGluAlaAlaIy 383
2178 ACAGTGGAGGATACCCTGTGTCAAGGCCCAGTGG 2211
     :||||||||:::||||    :::|||
383 sGluTrpAsnIleGlulleIleAsnHisLeuTrpIleGluAspSerTyrA 400
2211 ....................................... 2211
400 laLysCysGluValGlnThrPheSerAsnProArgTyrGlnHisPhePro 416
2212 ........CTTGCGACATTCTTCTTGGGAAACTTT........GA 2240
                    :::||| ||| |||
417 ProArgThrAsnLeuGlyValGlnIleIleGlySerThrPhePheAspGluGl 433
2241 GGCACTGCGAGGCAGATTCAGTAT 2262
         ||||||||:::||  :::
433 nArgLeuArgGluValValTyrPheProGlyGlyGlyValGluValSerThrA 450
2263 ....................................... 2263
                                        AGTCCCTACACG 2274
450 laAlaLysSerLysArgGlnLysSerAlaAspAsnAspAlaAlaAsnLysAsnValTyr 466
2275 GCATTCAGTCTCTGCAGATCCATTTGCCCCTACCCACCAGCATTAGTTTAAA 2324
     ||||||                  :::|||
467 AlaSerAsnLeuAlaGluThrAspAlaGlyThrAlaAAsnLysSerSerPr 483
2325 TCTTTAGATGCTTGGAGAGTTCCCTTAAAAGTGTCTGCAGAGTTGTTGA 2374
                            :::|||:::|
483 oLeuGlyGlyLysThrAlaIleLysProProArgGlyAlaGlnAsnLysG 500
2375 TCAGTATAAAGACTACCTCCCAACTG................AAACAG 2406
     :::||| :::||  |||                         :::|||
500 lnAsnIleGlnThrProAlaLysThrArgThrThrArgGlyLysGlu 516
2407 AATGAAGTAGTAAAGTCCAGCCCTTCTTCCCAAAGA 2442
     |||:::||||:: |||:::||
517 AsnAspThrProSerValMetSerSerGlySerArgSerAlaLysAspGly 533
2443 .......GCCAGAATTGAAGACGTACCACCT 2467
                  :::|||             |||
533 sAlaLeuSerLysLysLeuHisLeuGlnLeuLeuAlaProAlaAspIleAlaLeuThrArgGluL 550
2468 CCACTCAAAAAGCTAACTCCAGAAATTGACCCCTTTTGTCCTTTTCACTGGA 2317
     ||||||  |||:::||        :::||
550 ysGlnLysLysArgThrAlaLysSerAspProGlyProTrpGlyLysArgAla 566
2518 TTCGAGCCCTGTCCCAGGTTCAACAGTATAATTAAGAAGCTCACATTCTT.. 2565
     ||||||||        |||:::|||
567 AlaAspGlnIleAlaSerpArgAspArgAlaLysCysGluSerSerProAlaGl 583
2566 ......GGTGGAGAGGTTGGGGAGTCTGCACAGNAG 2595
            ||||||||         :::|||
583 uAspGlyGlyLeuGlnAspGluThrGlnLysAspLysSerArgProAlaLysLysA 600
2595 ..................................... 2595
600 lAArgAspGlyGlyMetArgGlnMetArgIleLeuCysLeuThrGlyTyrLys 616
2595 ..................................... 2595
617 ArgTrpValAsnAspLysMetArgGluGlyAlaAspsArgIleValGlnAs 633
```

```
2596 ..........TGCACACACCTCATTCGCCACGAAAGTGACCTCGCCACGCTGA 2635
              :::||| :::|  |||||:::|||  ::||||||:::|| |
633 pAsnValProCysAspTyrLeuAlaLeuAlaProArgMetValArgThrMetLL 650
2636 AGTTCCTCACGCGCGATTTCTGCGTGCGCTGGAGGCACACTAGTGACGCCAGAGTGG 2685
     ||||||||              ||||||:::|||
650 ysPheLeuArgCysLeuAlaArgGlyProAspIIeIIeSerSerAspTyr 666
2686 CTGGAAGAATGCTTCAGGTGTCAGAAGAGTTCATTGAAGACAGAGAACTACAT 2735
     :::|||  ||||||:::|||||
667 ValThrAlaCysGluValGluAlaGlyIleLysValValLeuProProLysGly 683
2736 TCTCCGAGATGCCTGAGGCGAGGAAGTACTTTCTCTTTCAGCCTTGGAAGAAT 2785
     ||||||  |||||||:::||||
683 uLeuValAspLysGluSerGlyUAspArgPheGlyValThrLeuGlnThrA 700
2786 CCTTAAAACGGGCACACGTTTCTCCACTCTTTAAGGCAAAATATTTTTTAC 2835
     :::||||||||:::||||
700 laIleSerArgAlaArgGlAlaAsn.........ArgGlyArgLeuLeuTrp 713
2836 ATCACACCTGGAAATCTGCCCA.........AGTCTTTCCACTAT 2870
     :::||| :::||       :::|||
714 GlyValProValPheCysThrGluGluIIeLeuLysAsnGlyValGlnSerTyr 730
2871 GAAGGCAAATCCTAGAGTGTGCAGGA 2895
     |||:::|||||        :::|||
730 rGlnThrIleAlaGluAlaAsnGly 738
```

Quality: 223.50 Length: 303
Ratio: 1.380 Gaps: 9
Percent Similarity: 53.465 Percent Identity: 27.723

alignment_block:
US-09-664-641-10 x RGHY55

Align seg 1/1 to: RGHY55 from: 1 to: 905

```
817 CCTCGTCTGATTATTATGAAGAGGAAGAGGAAGAGGAGGA
      |||:||      :|||:::::||::::||::  :::::
 42 ProGlnGlnIleLeuGlnMetIleGlnIleArgHisGlnIleLeuLeuArgSe 58

867 CGAACTAGAAAAATGAGGAACAAAGAGATTCTCAGAATCAGCGGTAGTACAGATG 866
      :::||::||:||:||||||||||| :|||||||||:||||||||||   ::
 58 rArgLeuGlnLeuGlnGlnGlnGln                          .G 69

917 AGAAGTCAGCCTCCAGCGTCCTTGCTTCAGGTGACCAG 966
      ::  ||| |||  ||||||||| |||
 69 InThrSerProProProGlnThrHisGlnSerProProProGln 85

967 CAGTTTTCACCTAAATCCAACACTGAAAAANTCTAAAGGGGAATTAATGTT 1016
      |||  ||| :||:|||| ||||:::|||   :|||||
 86 GlnSerGlnProProIleAlaAsnGlnSerAla.. 95

1017 TCATGATTCATCACCCGCAAAAACAGGAGCAGAAATTTAAACT 1066
      :||:|||||:|:  ::|||||
 96 ..ThrSerThrProProProProProAlaProHisAsnLeuHis. 109

1067 GGACCCCCGCCGAAGTCCCACAGTTAGCTGCACCAAAACGCAGCGTGCCT 1116
      |||||||||:::
 110 ..........ProGlnIle.. 112

1117 CAGGGAAAGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
      |||:||      :||||
 113 ..GlyGlnValPro.. 120

1167 CCCAGGTAACACATTTTGCCCCCTCGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
      :||  ||| |||
 120 aProIleAsn..LeuProGlnIleLeuProLeuAlaThrG 136

1217 GACAAAAACCTCCACAAAGTTCTGAAAGATCAGAAAATGATAGCTACCTGGAGT 1266
      |||:||      :|||||
 136 InGlnLysValLeuAsnLysLeuArgGlnIleAlaIleAlaLysAsnAsn 152

1267 CCAGGTGTACGGACACTGAGGAATATTACTAATAATGTGACATCAGCA 1316
      ::  ||| |||
 153 ProGlnValValAsnAlaIleThrValAlaGln.. 165

1317 GATGAAACCGGCACTCAAAATGTAGGACATATCTTACACAGTCTTTCAGCAC 1366
      ||||:||      :|||||
 165 nValGlnArgGlnIleLeuGluGlnIleLysGlyGlnIleThrValAlaGln 182

1367 CTAGCAAAAATTTACAACAGCGACCACGTCAAATCCACCGCCAGCAGACATACA 1416
      |||||      ::||||
 182 InLeuGlnGlnIleLeuGlnLeuValGlnIleGlnIleGlnIle 198

1417 AATGCCAATGCAGTGCGTGTTTACCAAGTGAAAGTGACTCCAGACACA 1466
      ||||:||      :||||
 199 LeuArgAsnGlnIleGlnArgGlnInPheArgHisIleSVa 215

1467 CATGCTACACGCAGCAGCACCCAGCAGCAGCAGCACCCGC 1516
      |||:||:|||||||||||||||||:||
 215 lGlnIleGlnIleGlnLysGlnGlnIleGlnGlnGlnGln.. 230

1517 TTTTACACCTTCAGCCCCCACAGATAATGCCAGTGTTCCAGCACGCACCAGCAG 1566
      |||  ||| |||||||||||||
 231 ..HisGlnGlnIleGlnIleGlnIleGlnIleGlnGlnGln 245

1567 CAGCAGATCTCTGCAGCAACCTTACCCCCAGCACGCGGCGCATCCATTTTC 1616
      |||:||      :||||
 246 GlnGlnGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 262
```

seq_name: pir2:S69205

seq_documentation_block:
stripe a/b protein - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C;Accession: S69205
R;Frommer, G.; Vorbrueggen, G.; Pasca, G.; Jaeckle, H.; Volk, T.
EMBO J. 15, 1642-1649, 1996
A;Title: Epidermal egr-like zinc finger protein of Drosophila participates in myotube
A;Reference number: S69205; MUID:96203082
A;Accession: S69205
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA <FRO>
A;Residues: 1-1180 <FRO>
A;Cross-references: EMBL:U42403; NID:g1147788; PIDN:AAB02355.1; PID:g1147789
C;Keywords: alternative splicing

alignment_scores:
Quality: 220.50 Length: 650
Ratio: 0.817 Gaps: 26
Percent Similarity: 41.538 Percent Identity: 21.538

alignment_block:
US-09-664-641-10 x S69205

Align seg 1/1 to: S69205 from: 1 to: 1180

```
1617 ACAGCACCAGCAGCAGCAACAGCAGCCACCACCATCG.......CCTC 1657
      ||||||||||||||||||||| |||   ||| |||      |||
 262 nGlnGlnGlnGlnGlnGlnGlnIleProGlnSerGlnValProGl 279

1658 AGCAGCATCACCTTTTTGGACATGATACCAGCAGTGGAGGATTCCAGAGA 1707
      ||  :::::|| ||||:|::|||   |||   ||||||||||| |||
 279 lnValArgSerMetSerGlyGlnProProThrAsnValGlnProThrIle 295

1708 GGCTTCTTA 1716
      |||
 296 GlyGlnLeu 298

865 GAGGAAGTAGAAAAATGAGGAACAACGAGATTCTCAGAAnCGACGGTACTACAGA 914
      ||||||      :|||||||:|||
 443 GluGlnValGluAlaHisLysGln...... LeuIleGluAlaLeuProGl 457

915 TGAGAAGTCAAGCCCTGCCAGCTCTCAAGAGAAGGGTCCTCAGGTGACC 964
      :::||::||::|:|||
 457 yAspLeuAsnThrProValThrThrSerSerAspIleProSer. 471

965 AGCAGTTTCACCTAAATCCAACACTGAAAAATCT......AAAGGGAA 1008
      |||||      :|||| :::|||
 472 .PhePheGlyProThrThrValGlnGluProProIleThrGlySer 487

1009 TTAATGTTTGATGAT.................. 1023
      :::  :|||
 488 IleGluSerGluAspLeuSerLeuGluProGlnValIleSerValAlaSe 504

1024 ......TCTTCAGATTCATCACCGGAAAAACAGGAGAGAAATTTAAACT 1066
      ::::||||:::
 504 rProValLeuSerHisCysSerProLeuLysGluUAArgSer. 518

1067 GGACCCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGCGTGCCT 1116
      :::||||:||||||||||||||||:||||||||||||||||||:|
 519 .ThrPro.......ProAlaLeuAlaLeuVal.. 526

1117 CAGGGAAAGAGCCTGGGTTGATTAACTTGTGTGCCAATGTCCCACCCGT 1166
      ||||||      :|||
 527 ..LysGluGluSerSerAsnAsnSerCys.. 535

1167 CCCAGGTAACACATTTTGCCCCCTCGAGGTCCGGGGTAATTTAATGGCTGCTG 1216
      |||:||      :|||
 536 ..AsnMetTyrProGlnHisAsnAsnAsnAsn.. 545
```

```
1217 GACAAAAACCTCCAAAGTTCTGAAAAGATCAGAAAATGAATAGACTACCTGGAGT 1266
          :::::|||                                     :::|||
 546 .........................AsnAsnAsnAsnAsnThrThrSerSer 554

1267 CCAGCTGTACGGACGACTGAGGAATAATTACTAATAATGCTGACATTCAGCA 1316
          |||:|||||                      ::|||
 555 SerThrThrThrThrSerLysGlnThrThrSerGluSerAsnThrGluCy 571

1317 GATGAACGGCCCATCAAATGTAGCACATATCTTACAGACTCTTCAGCAC 1366
          ::|||:|||||                      :::|||
 571 sValGlySerProGlyAsnHisThrGlnSerHisGln............... 583

1367 CTACGAAAAATTAGAACAGCAGGTGAAATCACCAGCCAGCAGGACATACA 1416
          ::|||||||||:::|||:||||                 :::
 584 .........GlnGlnGlnThrGlnGluValGlnHisAsnThrSerSerSer 597

1417 AATGCCAATGCCTGCTTGTTAGCCCAAGTGAAAGTGACTCCAGAGACACA 1466
          ||||||||||                      |||
 598 AsnSerAsnCys.......................HisHi 603

1467 CATGCTACAGCAGCAGCAGCAGCAGCCCCAGCAG..AGCAG 1503
          ||||||||||  |||||||||                |||
 603 sSerHisGlnGlnGlnGlnGlnGlnGlnGlnHisMetSerSerP 620

1504 .......CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG... 1536
          ::|||||||||||||                    |||
 620 roGlnGlnGlnTyrGlnGlnHisGlnIleLeuHisGlnGlnGlnPhe 636

1537 ................................CAGAT 1541
                                            |||
 637 GlyTyrHisHisHisHisHisHisHisHisHisHisHisHisSerGlnIle 653

1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGTCTCACCAACCTTACC 1591
          ::|||||||||                      :::|||||
 653 uGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisG 670

1592 CCCAGCAGCCGCGCGCATCCATTTTCAGCAGCGAGCAACAGCAG 1641
          |||||||||||                :::|||||||||
 670 nGlnGlnProLeuHisGlnGlnGlnGlnGlnGlnGlnGlnGln 686

1642 CCACCACCATCGGCCTCAGCACCATCAGCTGTTTTGGACAATGATCCAGCAGT 1691
          |||||||||                      :::|||
 687 ...........GlnHisHisGlnHisTyrGlnGlnHisHisAlaIl 698

1692 GGAGATTCCAGAAGAAGGCTTCTTATTGGGATGTCTGTTGCAATTCCGG 1741
          :::|||                      :::
 698 eGln................................................ 699

1742 ATTATCCACAGAGCAGATGTCTGATAAGCAACTG...CTGGCCACCTGGAAA 1788
          :::|||||                      :::|||
 700 .........HisGlnGlnSerGlnGlnIleAlaSerLysIleSerTyrArg 713

1789 AGGATAAATCCAGGCACAATGGCGGCCACTGTTGACCCCACCTTCAGGAGTCG 1838
          |||:|||                      :::
 714 GlyIleProPheThrThrGlyAsnAlaMetAsnAlaLeuAlaLaAlaAl 730

1839 ATGCACGCACCTTCTCTGTGAGACTCAAGTCAGCACGCGGCCTATGCACAGG 1888
          :::|||                      |||
 730 aAlaAlaAlaIaGlnGlnHisGlnGlnGlnHisGlnHisGlnGlnG 747

1889 CAATAAGAGAAAGAAAGAGAGATGTGTTACTGCACACTGGTTAAACACAGTC 1938
          :::|||||                      |||
 747 InLeuProSerProGlnLeuGlyValLeuAla................... 757

1939 TTAAAAAAAAAAAAAATGTACCCCCGCCCACCAGCCCTTCACTTCCCAGT 1988
          |||||||||                      |||
 758 .........GlyProMetSerProProSerAsnSerLeuGlyAsnSerTr 771

1989 GGCCTTCCCACCAGGAGGAAAGCCATGTTCACAG...................... 2022
          :::|||||  |||
 771 pGlyLeuProSerProAspLysThrMetPheGlnProProLeuPheSerL 788

2023 ..................CATATTATTTCTGTG...................... 2037
 2023
```
（右欄 / right column）
```
 788 euProAlaHisTyrAlaThrMetGlnGlnGlnGlnGlnGlnGln
     :::::
2038 ........................ACTGGATTTGTTGATAGTGA 2057
                              ::::|||
 805 GlnGlnGlnGlnGlnGlnGlnGlnGlnAlaAlaGlyIleAlaLaAlaProSerPr 821

2058 CAGAGAATGACCTAAAAATTAATGGCTTATTTGGCAGGT 2101
          |||||||                      GCCAAAT
 821 oTyrAspAspGlyArgGlnAlaAlaAlaAlaAlaAlaAlaGlnHisAlaGluL 838

2102 ATACGGGTTATCTATGCCCGCACAACACAGTCCTCATCTGTAAAGAACCA 2151
          :::|||                      :::|||
 838 euLeuGly..LeuThrMetAspCysThrProLeuLeuLeuLysGlnPro 853

2152 ACTGGTTTTAAAAGTATGAAAAAGTGGAGGATACCCTGTGTCAA 2201
          ::::|||                      |||
 854 ProPro...SerTyrAlaGlyAlaSer.....................AlaGl 863

2202 CGCCCAGTGGGCTTGGCGACATTCTTCTGGGAAAACTTTGAGGGACTCAGGC 2251
          ::|||||||||                      :::
 863 yPheProGlyLeuGlyAspLeuHisSerSerHisGluGlnGlnGlnGlnGln 880

2252 AGATTCAGTATAGTCGGCTACACGGCATTCAGTCTGCAGGATCCATTGCC 2301
          :::|||                      :::|||
 880 InGlnGlnTyr............................................ 883

2302 CCTACCCAGCCAPTTAGTTTTTAAAATCTTTTAGATCGTTGGAGAGTTCCCTT 2351
          ||  |||
 883 ....................................................... 883

2352 AAAAGTGTCTGCGAGAGTTGTTGATGAGTATAAGACTACCTCCCAAACTG. 2400
          :::|||                      |||
 883 ....................................................... 883

2401 .............................AAACAGAATGAAGTAGCT 2418
                                        :::|||
 884 ........................ValArgSerGlnProGlySerTyrG 891

2419 AATGTCCAGCCTGTTCTTCCCAAAGAGCCCAGAAATTGAAGAGACGTACCACCTCC 2468
          :::|||                      :::|||
 891 lnArgLeuAspSerProAlaAspTyrAlaGlyGlnGlnGlnGlnGlnValGln 907

2469 CACTAAAAAAGGCTAACT 2484
          |||:|||                      :::
 908 GlnValGlnGlnGlnGlnGlnGlnGlnThrLeuValLeuProGlyPr 924

2485 ..............CCAGAATTGACCCCTTTTGTGCTTTTCACTGGATTC 2520
          :::|||                      :::|||
 924 oThrSerSerAlaSerSerAsnAlaLeuAlaLeuGlyValLeuIlePro 941

 941 ysGlnGluAsnTyrProAspMetGlnProSerSerAsnGlyThrGlyTyr 957
```

```
seq_name: pir2:T14577

seq_documentation_block:
protein kinase YakA (EC 2.7.1.-) - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14577
R;Kuspa, A.; Lu, S.; Souza, G.M.
submitted to the EMBL Data Library, January 1998
A;Description: YakA, a protein kinase required for the growth to development transit
A;Reference number: 218146
A;Accession: T14577
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1457 <KUS>
A;Cross-references: EMBL:AF045453; NID:g2854116; PID:g2854117; PIDN:AAC02554.1
A;Genetics:
A;Gene: YakA
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific prote

alignment_scores:
```

```
              Quality:  219.50          Length:      397
                Ratio:    1.155           Gaps:        15
Percent Similarity:  47.859   Percent Identity:  23.929

alignment_block:
US-09-664-641-10 x T14577

Align seq 1/1  to: T14577  from: 1  to: 1457
```

```
 835 GAAGAGGAGAAGAGGAGGAGGAGGAGGAGAAGTAGAAAATGAGGA 884
         ::::::::::::::::::::::::::::::::::::::
                                              .CCTGCCAGCTCT 939
 928 ..............................................
         :::::::::|| :::||||
 961 ..............................................
                                                .GACCAGC 967
 968 AGTTTTCACCTAAAATCCAACACT 996
 997 TCTAAAGGGGAATTAATAGTTTGATGANTCTTCAGANTCANTCACCGGAAAA 1046
1047 ACAGGAGAGA 1078
```

(sequence alignment data — rotated, partially legible)

```
alignment_scores:
              Quality:  213.00          Length:      324
                Ratio:    1.275           Gaps:        11
Percent Similarity:  51.543   Percent Identity:  26.543
```

alignment_block:
US-09-664-641-10 x S37876

Align seq 1/1   to: S37876   from: 1   to: 738

```
  835 GAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGAGGAAGATAGAAAAATGAGGA  884
      ::|||:::|||::||||||||||||||::::|||||||::|||
  199 GlULysGlUGlUGlUGlnGlnGlnGlUGlUGlyHisAsnAsnLysGl        215

  885 ACAAGAATTCTCAGAATGAG..............................      903
      :::||| | ||||
  215 uGluHisLysGlnIleGluGlnProSerLeuSerSerLysLysThrThrS     232

  904 .........GGTAGTACAGATGAGAAGTCAAGACCCTGCCAGTCTCAA        942
      :::||| ::|||||  ||| |||:::
  232 erArgThrSerAlaSerGlnProLysLysMetSerTrpAlaAlaIleAla     248

  943 GAAGGGTGCTTCAGGTGACCACCAGTTT...TCACCTAAATCCAACAC        989
      ::|||::|||:::||:::                 |||||| :::
  249 ThrProLysProLysAlaValLysLysThrGluSerProLeuGluAsnVa     265

  990 TGAAAAATCTAAAGGGGAATTAAATGTTTGATGAATTCTTCAGATTCATCAC    1039
      :::||||||||| |:::
  265 lAlaGluLeuGlyLysGluIleSerAsplleLysLysLysAspGlnLysS     282

 1040 CGGAAAAACAGGAGAGAAATTTAAACTGACCCGCGCCGAAGTCCCACAG      1089
      :::|||||||||||||| |||
  282 erGluAlaSerGluGlyLysValAsn                            292

 1090 TTAGCTGCACCAAAACCCAGCGCTGCCTCAGGGAAAGGACCCTGGTTGAT    1139
      ::|||::|||:::||:::                 |||||||:::
  293 GluThrSerAlaGlyGluLysThrAlaGluProSerGluGl            309

 1140 TAACTTGTGTGCCAATGTCCCACCCGTCCCCAGGTAACATTTGCCCCTG      1189
      ::|||:::||||:::
  309 uAsnGlu...AspArgValProGluValAspGlyGlu                  321

 1190 AGGTCCGGGGTAATTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAA     1239
      |||||:::                           |||
  321 luValGln                                               327

 1240 AGATCAGAAATGATAGCTACCTCGAGTCCAGCTCGTACGGACACTGAGGAA   1289
      :::||||||||:::|||  ||:::
  328 LysLysGlyGluLysValLysGlnThrAlaGluGluLeuGluGl          344

 1290 TATTACTAATAAT........GCTGACATTCAGCGAGATGAACCGGCCATCAA 1333
      :::|||                      :::
  344 nGluGlnAspAsnAsnValAlaAlaIleProGluGluValAlaThrValGluG  361

 1334 ATGTAGCACAATATCTTACAGACTCTTTCAGCGACTTACAAAAAAITTAGAA 1383
      :::||| ||| :::                    :::
  361 luLysValGluIleSerAlaValIleSerGluThrProGluGlnAla        377

 1384 CAGCAGGTTGAATCACAGCCAGCCAGCAATACAAATGCCAATGCCACTGCT   1433
      :::||||||||||||||| :::
  378 AsnThrValProGlnProGlnGln                               385

 1434 GTTTAGCCAAGTGAAAGTGACTCCAGAGACACACATGCTACACGACGAGC   1483
      |||                         |||
  386 .....GlnSerGlnProGlnProGlnGlnProGlnProGlnProG        400

 1484 AGCAGGCCCCACGAG.....CAGCAGCAGCCCCGGTTTACACCTTCAG     1530
      ||||||||| :::                      |||||
  400 lnGlnProGlnGlnProGlnGlnProGlnGlnProGlnGlnProGlnGln     416

 1531 CCCAGCAGATAATCCAGTCCAGCCAGCCAGCCAGCCAGCAG            1569
      |||||||||||| :::
  417 ProGlnGlnProGlnLeuGlnGlnGlnGlnGlnProValGl             433

 1570 ...................CAGATCTCAGCGAACCTTACCCC             1594
      |||
  433 nAlaGlnAlaGlnAlaGlnProGluGluVaLGlnLeuSerGlnAsnThyrThrG  450
```

seq_name: pir2:T13167

alignment_block:
US-09-664-641-10 x T13167

alignment_scores:
    Quality: 210.50   Length: 334
Percent Ratio: 1.126   Gaps: 15
Percent Similarity: 55.988   Percent Identity: 26.347

Align seq 1/1   to: T13167   from: 1   to: 1010

```
 1595 AGCAGCCCCGGCCATCCAATCCATTTTCACAGCAG...................   1623
      |||||                    :::|||||||||||
  450 lnGlnGlnGlnGlnGlnGlnTyrAlaGlnAlaGlnGlnHisGlnGlnGlnGln   466

 1624 ....................CAGCAGCAGCAACAGCAGCCAC.            1646
      :::                   ||||||||||||||::::|||
  467 TyrLeuSerGlnGlnGlnGlnTyrAlaGlnGlnGlnGlnGlnHisProGl     483

 1647 ACCATCGCCTCAGCAGCCATCAG                               1668
      ||::                ||||||  |||
  483 nProGlnSerGlnGlnProGln  490

  769 CTGGATTGCGTATCAGAGGAAACCAAA....AAGGACGAAGCAITTTATCA     815
      ||||||                    :::|||||| ||||||||:::
  411 LeuAspLysSerAsnHisLysGlnLysIleLysAspAsnSerIleVaLhrh  378

  816 TCCTCGGTCCGTATTATTATGAAGAGGAAGAGGAAGAGGAAGAGGAAG     865
      :::|||                ||| :::
  378 rGluMetValIleGluProLysAlaGluThrAspAspAspAlaHisAspG     395

  866 AGGAAGTAGAAAAATGAGGAAGAAGATTCTCGAATGAGGGTAGTACAGAT   915
      |||::::|||||||                 ||||||||
  395 luAsnValGluAspLeuThrLeuAsp...GluGluAspMetThrMetGlu   410

  916 GAGAAGTCAAGACCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCA  965
      |||:::|||||||||||                   :::
  411 GluLeuAspGlnAlaLeuGlyThrSerGlnLysGluGlySerSerGl     427

  966 GCAGTTTTCA......CCTAAAATCCAACACTGAAAAATCTAAAGGGGAAT   1009
      |:::||                  :::|||||||||||||||:::
  427 nAlaTyrAlaThrProGlnHisAspAspAspArgSerGlnAspGluLeuGlyLeuM  444

 1010 TAAATGTTTGATGAATTCTTCAGATTCATCACCGGAA.....AAACAGGAG  1053
      |||||::::                       ::::|||
  444 etAlaAlaGlnAspAlaGlnArgAspProGlnAsppLeuGluLysGlnAsp  460

 1054 AGAAATTTAAACTGACCCCGCGCC...........GAAGTCCCACAG.      1089
      |||                       :::||| :::
  461 LysSerGlyGluGlnThrGluGlyAlaHisAspGluPheGlIleLeuAspAspCy  477
```

```
1090 ...TTAGCTCCAGCAAAACCGACGCTGCCTCAGGAAAGCTGGGTt 1135
          ||| ||:::||  :|  |
477 sLeuLeuGluSerAsnAspIleValIleThrGlnAsnLysAsp...... 491

1136 TGATTAACTTGTGTTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCC 1185
          ||| 
492 ...............................GlyPheValLeuHis 496

1186 CCTGAGGTCCGGGTTAATTTAATGGCTGCTGGGACAAAACCTCCAAAGTTC 1235
     ||||:||||| ||||||
497 ValLysLysLeuGlyAsnIleThrAlaAlaLysMetGlySerThrThrAs 513

1236 TGAAAGATCAGAAATGATAGCTACCTACCGCTACGGGACTACGGA\ACTGA 1285
     :||||:|:: |||||||
513 pGluGlnAlaAlaAlaValThrValThrGlyTyrProAlaGlyYGlnProThrG 530

1286 GGAAATATTACTAATAAATGCTGACAATTCAGCAGATGAACCGGCCATCAAAAT 1335
     ::||||| ||||||||||
530 ThrThrIleThrGluLeu......LeuSerGlnAlaAlaHisSerSerAsp 544

1336 GTAGCACATATCTTACAGACTCTT...TCAGCACCTACGAAA 1374
     :|||||||| |||:|||||| |||
545 ValLysProThrLeuThrThrLeuThrAsnThrProIleLysLeuProSe 561

1375 AAYTTAGAACAGCAAGGTG............ 1392
     ||| ||:||| :::
561 rSerGluCysGluLeuIleLeuAsnIleLysLysIleIleProSerSerThrT 578

1393 ....AATCAGCACAGCCACAGGGACATACAAAATGCC 1422
          ::|| |||||||||||
578 hrIleuAla\ThrHisGlyCysGlySerSerSerSerGlyThrThrThrT 594

1423 AATGCAGTGCGTGTTAGCCAA...GTGAAAGTGACTCCAGAGAC 1463
     :::||||||||
595 ThrThrThrIleIleHisProHisHisIleIleLeuHisValThrGlnGluTh 611

1464 A..........CACATCCTACAGCAGCAGCAGGCCCCAGCCAGCCAGCAGC 1504
     A          |||| ||
611 rGlnGlnGlnHisGlnGlnGlnGlnGlnGlnHisGlnGlnGlnGlnGlnG 628

1505 AGCAGCACCCGGGTTTTACACCTTCAGCCCCAGCAGANTAATGCAGCGTCCAG 1554
     ||||| ||||||||||||
628 lnGlnGlnGln.......GlnGlnGlnHisHisGlnGlnGln 640

1555 CAGCAGCAGCAGCAGCAGAGATCTCTCAGCAACCTTACCCCAGCAGCCGCC 1604
     ||||:: |||||
641 GlnGlnGlnGlnGlnLeuHisLeuHisIleGluGluValProGlnSerSerHi 657

1605 CCATCCATTTTCACGACGACGACGACGACAACAGCAGCCACCACCATCGC 1654
     ||| |||||||||||||:::
657 sHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln.HisHisHisGln 673
```

seq_name: plr2:S44898

seq_documentation_block:
ZK1236.3 protein - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
A;Reference number: S44622
A;Accession: S44898
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1000 <FAV>
C;Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 23-Mar-2001
C;Accession: S44898
A;Cross-references: EMBL:L13200; NID:g289748; PID:g289754
C;Genetics:
A;Genetics:
A;Introns: 56/3; 86/2; 189/3; 243/2; 669/3; 808/3; 984/3
A;Description: Sequence of the C. elegans cosmid ZK1236.
R;Favello, A.D.
submitted to the EMBL Data library, May 1993

alignment_scores:

alignment_block:
US-09-664-641-10:

Align seq 1/1 to: S44898 from: 1 to: 1000

Quality: 204.50		Length: 1006
Ratio: 0.492		Gaps: 47
Percent Similarity: 41.352	Percent Identity: 19.682	

```
139 AGGGCTACGGATCTCTCCGGCCCCTGCCCTTCAAATCACCTCTAGGCACAG 188
    |||  |||   |||  |||||||:||||||
104 ArgMetLysAspThrSerLeuProProLeuGluAsnProSerAsnGlyGl 120

189 GTTATTCAGCTGTTCTCAAGGCTGGAAAAGCCGAAGGAAGTTYCCTACAAATGC 238
    ||||:|||  ||| ||:|: ||||||||
120 yThrThrTyrLeuPheAsnLeuIle.......ThrValPheProSerIleA 134

239 ACTAGCCTCACACATAAATCTCAGAGGATGGGACAAATCCAGAGGTGGCAG 288
    |||||||:||| |||:||:|:|||||||
134 laLeuSerPheAsnTyrSerTrpArgGlyAspThrGlyProSerIleSer 150

289 AAGCTCGGGAAGTCTTTGACTTACCTGTTGTTAAAGCCTTCTTGGGTGATT 338
    :::|| ||||| |||||||||||:|||||
151 IlePheSerPheSerLeuSerValLeuPhePheSerLeuPhePheProGlnHi 167

339 CTGTCCCGTTCAGTGTCGGAACTCTCTTGCCCAGTAAAAYGGYTTYTCTCCCAGA 388
    ::||||| |||||||:||:|||||
167 sLysAsnIleCysAlaAlaArgAlaTrpCys................ 176

389 ATCATGTCAGATTTTTTTTGGAATCACTGCCTCGCCTTTCTCAGGGTGTTG 438
    |||  |||  |||:||||||||||:::
177 .....ArgProPheArgArgSerLeuSerPheSerLeuArgPheIle 190

439 ATACAAGCTG/AGCTCTTTGTTGGGAGTCTTCCAGAGCTCTCCAAGGAGA 488
    :::||||||    ||
191 MetSer..............SerGluProAlaSerSerSerThrGl 201

489 G....GTAGGGAAGGGAGCGTTGTCCAGCAGAAGTTGGGAAGCACAGAGATC 535
    |     :::||  |||||||||||:::
201 lnGlyProGlnPhe.......ValIleProArgAlaLeuSerArgAspTyrVal 217

536 ATCTGGCTTCTGTTCTGACCCGGTATTGATCAGGCCTCAGGCGCTCGTTGTTA 585
    |||  |||  |::|||  ||||||||:::
217 lnGlyProGlnPhe.......LysHisLysPheG 217

586 ATGTCGTGGGTGTGTCATCTGAAGACAGAGAAGTGCCCGTCGGGCTTGGGTTAC 635
    |  ||||||||||||:::
232 LeuAsnValAsnAsnGlnThrProGluSerTyrGlLysAlaLeuAsnAl 248

636 GTTCTATGGGGAGAGATTGCCAGCTTAACCCTCAATAAGAAATGCACGCATT 685
    |||||| |||||||||:||||||
248 aLysTyrGlyArgAspAspTyrIleThrLeu........CysLeuHisV 262

686 TGATTGTT..........CCAGACCAAAGGGGGGAAATAC 717
    |:                 ::|||||||
262 alThrSerLeuCysThrGluTyrGlyProLeuAspAspGlyProGluTyr 278

718 GAA......TGTGCTTTAAAGCGAGCAAGTATTAAAATTGTG 753
    |||           ||||||:::::
279 ValLeu....uCysAspSerArgAlaAlaArgValAspLysLeuIleGluAspIle 295

753 ........... 753

754 ...........ACTCCTGACTGGGTTCTGGATTCGGTA 780
    :|          |||||| ||||||
295 uValLysLeuLeuGluIleAspThrArgAspTyrValGlnLeuLeuHisG 312

781 TCAGAGAAAAACAAA.......AGGGACGAAGCAYTTTATCATCTCTG 821
    ||| |||        |||:|:|| ||||||:::
312 lyGlyLysArgLeuHisLeuGlnLysProAspAlaValLeuArgAspIle 328

329 AlaTyrLysGlnSerAsnGluGlySerAspLysPhePheLeuGlMetLy 345
```

822 TCTGATT...ATTTATGAAGAGGAAGAAG 847
      :|||||| ||||||||||||||||||
345 sLeuValProSerGluSerMetLysAlaLysIleMetLysGlnGluGluG 362

848 AGGAAAGAG......................................GAAGAGGAGGAGGAAGTA 873
      |||||||| :::::
362 luGluGlyLysAlaArgLysHisGlyGlnTyrGlnGlyGlnGluTyr 378

874 GAAAATGAGGAACAA..............GATTCTCAGAATGAGGGTAGTACAGA 914
      :::: ||||||||||::::::|||:::
379 HisGlnGlnHisGlnAlaMetAsnAspGlyGlnSerSerSerSerValPr 395

915 TGAGAAGTCAAGCCCTGCCGCAGCTCTCAAGAAGGTCT.............. 951
      ::: ||||||||||::||::
395 oSerThrSerSerProSerCysSerSerGluAlaAsnArgLysGluMetG 412

952 .............CCTTCAGTGACCACCAGCAGTTT............. 972
                   |||||||||||||||||||
412 luThrValArgGlyProAlaGlyProSerGluLeuMetArgAlaIleAsn 428

973 ....TCACCTAAATC 983
                              :: |||
429 AlaProValAlaProAlaProValValIleLysIleGluThrProValAl 445

984 CAACACTGAAAAATCTAAAGGGGAATTAATGTTGATGAATTCTTCAGATT 1033
      |||||||| ||||||||||
445 aLeuProGluGluAspGlu...ThrLeuMetAspAspAspGluMetProS 461

1034 CA... 1035
      ||
461 erLeuThrValGluAlaProSerGluGluAlaSerPheGluLeuGln 477

1036 ..TCACCCGAAAAACACGAGAGAAATTAAACTGG........ACCCCGC 1076
      |||||| ::: |||
478 ProSerProGlnValProGlnAlaSerIleGluGlyProSerGlnGlnGl 494

1077 CGAAGTCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCTCAGGGAAAGG 1126
      |||||| :::||:::
494 nGlnIleProGlyThrSerGlnInLysArgGlnValAlaAArgGlySerA 511

1127 AGCCTGGGTTGATTTAACTGTGTCCCAATGTCCCACCCGTTCCCAGGTAAC 1176
      ::::::||:::||| ::::||:::|||
511 rgThrAsnMetIleSerTyr...HisAspLeuProGlyThrGlyAsn 526

1177 ATTTTGCCCCTGCAGCTCCGCGGGGTAATTTAATGGGTGCTGGACAAAACT 1226
      ||||||| :||
527 ..AlaProPro...............................MetAlaCysProGlnValTh 536

1227 CCAAAGTTCTGAAGAATCAGAAATGATAGCTACCTGGAGTCCAGCTGTAC 1276
      |||||| :::
536 rLeuLysLeuGlnLysAsn.....................ValProPheGluAlaLysIleA 550

1277 CGACACTCAGCGGAATTAATACTAATAAT..........GCTGACATTCAGCACATG 1320
      ||:::::||::: ||||||||||
550 rgAlaValAlaGlyTyrThrArgLysProIleSerGluValGlnLysMet 566

1321 AACCGGCCATCAAAATCTAGCACATATCTACAGACTCTTTCAGCACCTAC 1370
      |||||| |||:::|
567 ..ArgProSerAspLeuGluSerIlePheHisSerIle........ 578

1371 GAAAAATTTAGAACAGCAGGTGAATCACAGCCAGGACAATACAAATG 1420
      ||||||| |||
578 ....MetAlaCysProGlnValTh 578

1421 CCAATGCACTCGTGTTTAGCCAAGTGACTCCAGAGACACACATG 1470
      ::: |||
579 ....CysIleAlaSerValGlnArgGlyIleLysArgArgAsnGluLeu 592

1471 CTACAGCAGCAGCACCAGCCCCAGCAGCCAGCAGCAGCACCGGTTTT 1520
      ::||||||||| |||:||
593 ValGlnGlnGlnLeuGlnIleAsnAlaGlnSerCysLysSerProThrMe 609

1521 ACACTTCAGCCC........................CAGCAGATAATG 1545
      ::||||||||| :::
609 tThrMetAsnLysLysPheThrLeuAlaLysAlaThrArgGlnValGlnA 626

609 .......CAGCTCCAGCAGGACAGAGACAGAGATCTTCAGCAACCTTACCCC 626
        :::::: ||||||
609 tThrMetAsnLysLysPheThrLeuAlaLysAlaThrArgGlnValGlnA 626

1546 ..CAGCTCCAGCAGGACAGAGACAGAGATCTTCAGCAACCTTACCCC 1593
      :::::::||::: ||||||||:::
626 snGlnIleGluLysIleAspAspArgGluGlnIleIleLeuProGlnTyrMet 642

1594 CAGCAGCCGCCCCCATCCATTTCACACGCAGCAGCAACAGCAGCC 1643
      :::::::||||||| |||:::
642 snGlnMetProProMetPro........ProGlnGlyGlnGlnArgLeuPr 656

1644 ACCACCATCG...............CCTCAGCAGCAATCAGCTTT 1672
      |||||| ||||||||::::::
643 AsnMetProProMetPro........ProGlnGlyGlnGlnArgLeuPr 656

1673 TTGGACATGATCCAGCAGTGGGAGATTCCAGAAAGAGGCTTCTTATTGGGA 1722
      ::: ||||||||
656 OProProAlaTyrProProGlyIleIleLeuProGlnGlnAsnArgGlnG 673

656 OProProAlaTyrProProGlyIleIleLeuProGlnGlnAsnArgGlnG 673

1723 TGTGTGTTTGCAATTCGGATTATATCGAGCGAGAGTGCTGATAAGGAAACT 1772
      ::: |||
673 lnGlyValProGlnInPheGlnInArgSerProGln...PheMetIleGly 688

1823 CCACCTTCACCAGTTGGATCCCAGCGCCACCTTCTCTGAGAGTCAAGTCAGC 1872
      ||::: |::
702 roAsnSerAsnGlnArg...AlaArgIleLeuAsnThrSerSerValGln 717

1873 AGCCGGTATGCACAGGCAATAAGAGGAAAGAAAGAGATGTGGTTACTGCACA 1922
      |||||| |||
717 ..AlaArgIleLeuAsnThrSerSerValGln 717

718 ::::: |||
718 Pro.......SerGluValArgAsnArgLeuValLysIleGluAlaMe 732

1923 CTGGTTAAACACAGTCTTA..................AAAAAAAAAAAGGTAC 1960
      :::: |||
732 tAlaMetAsnMetAlaGlnLeuAsnProAlaGlnProProProGlnLeuP 749

1961 CGCCGACACCGAGCCCCTTCACTTCCCAGTGGCCTTC.........CCACCAGGA 2004
      ||||||| |::
749 roProHisArgAlaLeuGlnGlyLeuGlnPheLeuArgProGlyAla 765

2005 GGAAAGCCATGTTCACAGCAATATATTATTTCTGTGACTGGATTTGTTGATAG 2054
      ||||||| |||
765 ..AlaArgIleLeuAsnThrSerSerValGln 765

2055 TGACAGAGATGACCTAAAATTAATGGCTTATTTGGCAGGTGCCAAAATATA 2104
      ::: |||
770 ....SerGluValArgAsnArgLeuValLysIleGluAlaMe 770

2105 CGGGTTATCTATGCGCAGCAAACAACAGTCCTCATCTGTAAAGAACCAACT 2154
      || :::
778 hr..................TyrAsnAsnThrTyrValThrValAlaSerProAla 790

2155 GGTTTA............................AAGTATGAAAAAGC 2174
      ::: ||||||||::
791 ThrLeuThrAsnSerIleIleGluProTrpHisProProTyrGluLysSe 807

2175 C.............................AAAGAGTGGGAGGA 2188
                                     :::||:|||:::|||:
807 rGlyArgLeuAsnValSerAsnThrIleLysAlaIleAsnGlnTyrArgL 824

2189 TACCCTGTCTCAACGCCCAG.........................2208
      ::: |||
824 euLeuCysAsnSerArgIleAlaAspProAlaSerPheLeuGluPheTyr 840

2209 TGGCTTGGCGACAATTCTTCTGGAGAAACTTTGAGGGACTGAGGAGAATTCA 2258
      ::||||||||| :::
841 PheLeuGlyAspProMetProHisPheAsnLysIleLeuSerIleAlaAs 857

2259 GTATAGTCGCTACACGGCATTCAGTCTGCAGGATCCA..................2295
      :|||:::   :::
          ||| 2295

```
857 pTyrAsnMetTyrLeuSerArgArgArgCysAspGluAlaAspValLysI 874
        ||| ||||||||||
2296 TTTGCCCTACCCAGCACATTTAGTTTTAAAATCTTTTAGAT... 2334

2334 .......................................... 2334
874 leHisArgMetSerHisSerAspGlnLeuGlnLeuTyrLeuGlnLeu 890
        ||| |||||||::
2334 .......................................... 2334

891 GlnSerAspGluSerAspAsnValGlyLeuTrpLysThrPheTyrArgIleMe 907
        |||      |||
2335 .GCTTGGAGAGTTCCCTTA............................. 2352

907 tCInTrpAspLeuProLeuAsnAsnGluPheProArgIleLeuLeuProS 924
        ||||:|:
2353 ..................AAAGTGTCTGCA 2364

924 erSerLeuAspIleGlyArgProValValAspArgArgLysLysSerIle 940
        |||||||||||:
2365 GAGTTGTTGCATAGT............................ ATAAGACTACCTCCCAA 2396

941 AspGlnValMetLeuAsnHisIleHisArgMetHisSerGlnArgProPro 957
        |||:        |||||||
2397 ACTGAAACAGAATGAA.......................... ATAAGACTACCTCCCAA 2396

957 rMetGlyAsnSerSerThrSerSerGluAlaSerSerThrSerProThrA 974
        ||||||||||
2413 .GTAGCTAATGTCCAGCCTTCTCCCAAAAGACCACAGAATTGAA 2454

974 snAlaAlaThrAlaThrSerSerProAlaSerAsnArgProThrThrSer 990
        :::   ||||||||| |||
2455 CACGTACCACGTCCCACT 2472

991 ThrAlaGlnProProThr 996
        |||||||||)
2455 CACGTACCACGTCCCACT 2472
```

seq_name: pir2:A38346

seq_documentation_block:
ultra-high-sulfur keratin 1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 24-Sep-1999
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin o
A;Reference number: A38346; MUID:91065960
A;Accession: A38346
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-230 <W00>
A;Cross-references: GB:M37759; NID:g2009961; PIDN:AAA40106.1; PID:g2009962
C;Superfamily: ultra-high-sulfur keratin

alignment_block:
US-09-664-641-10/rev x A38346

alignment_scores:
```
      Quality:  200.00          Length:    218
        Ratio:    2.151            Gaps:     21
Percent Similarity:  42.661    Percent Identity: 31.193
```

Align seg 1/1 to: A38346   from: 1 to: 230

```
1814 GTGCGGCCATGTGCCTGGATTATACCTTTCCAGTGCCCACACAGTTGC... 1767
        |||||| |||||    |||||:
  56 ValProValCysSer.................................CysSerSerCysGl 65

1766 ....TTATCAGACACATCTGCTCTGGATAATCCGCAAATTGCAAACACACATC 1721
        ::::  |||||||:      |||||||||  ||
  65 yGlyCysGlySerSerCysGlyGlyCysGly.................. 75

1720 CCAATAAGAAGCCTTCTTCTGGAAATCTCCACTGCTGGA..........TCA 1680
        |||                                           |||
```

```
  76 ...........................SerCysGlySerSerCysGlyGlyCysGlySer 87
1679 TGTCCAAAAAGCTGATGGCTGC........................TCAGGGGATGG 1648
     |||||||||                                      ::::::
  88 CysGlyLysSerProValCysGlyCysGlyValProValCysSerCysSerSerCy 104
1647 TGTGGCTGC.......TGTTGCTGC.......................TGCT 1619
     |||||||||       |||||||||                        ||||
 104 sGlyGlyCysGlyLysSerProCysCysCysGlnSerSerCysGlyCysLysProCysC 121
1618 GTGAAAATGGATGCGGC...GGCTGCTGCTGGGGGTAAGGTTGCTGAGAGATC 1572
     |:::||||||||||||||   ||||||||||||||||||||||||||||||
 121 ysSerSerGlyCysGlySerSerCysCysGlnSerSerCysCysLysSerPro 137
1571 TGCTGCTGC..........................TGCTGCTGCTGGAGC. 1548
     |||||||||                           |||||||||||||||
 138 CysCysCysGlnSerSerCysGlyLysSerProCysCysCysGlnSerSerCy 154
1547 .TGCATTATTCTGCTGCGGGGCTCGAAGGTGTAAAACGCGGTGC.......... 1509
     |||||||||||||||||                              :::
 154 sCysLysSerProCys..........................CysSerSerGlyCysGlySerSerC 167
1508 ...........................TGCTGCTGC 1494
                                 |||||||||
 167 ysCysGlnSerSerCysGlyLysSerProCysCysCysGlnSerSerCysCys 183
1493 TGGGCCTGCTGCTGC........TGCTGTAGCATGTGCGTGTCTCTGGAGT 1453
     |||||||||||||||        |||||||||||||
 184 LysProCysProCysCysGlnSerSerCysGlySerSerGly.CysGlySerSerGly 196
1452 CACTTTCACTTGGCTAAAACAGCACACTGCAATTGGCATTGTAITGT...CCCT 1406
     ||||||||||                 :::   ||||||||||     ||||
 197 .........................CysCysGlnSerSerCysCysLysSerProC 206
1405 GCTGGCTGCTGTGAATTCACCTGCTGTTCTAAAATTTTTCGTAGGTGCTGAAAGA 1356
     ||     |||
 206 ys.........................CysCysGln............Serser 211
 206 ................CysCysGln................Serser 211
 212 CysCysLysSerProCysCysSerSerGly.CysGlySerSerCys......CysG 227
 227 ln 227
1305 AG 1304
     ||
 227 ln 227
```

seq_name: pir2:B38346

seq_documentation_block:
ultra-high-sulfur keratin 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 24-Sep-1999
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 266, 4024, 1991
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and ski
A;Reference number: A38660; MUID:91154184
A;Accession: A38660; B38346
A;Molecule type: DNA
A;Residues: 1-223 <W02>
A;Note: this is a correction
A;Cross-references: GB:M37760; NID:g2009963; PIDN:AAA40107.1; PID:g2009964
R;Wood, L.; Mills, M.; Hatzenbuhler, N.; Vogeli, G.
J. Biol. Chem. 265, 21375-21380, 1990
A;Title: Serine-rich ultra high sulfur protein gene expression in murine hair and ski
A;Reference number: A38346; MUID:91065960
A;Molecule type: DNA
A;Residues: 1-21; 'GGCGSGGGGGGGGSNGGGCGSSSCCKPVCC','22-40','GSS','44-45','G','47-48','S','50','GS
<W00>
A;Note: the sequence reported in this paper has been corrected. See A38660
C;Superfamily: ultra-high-sulfur keratin

Left column

```
alignment_scores:
     Quality:  198.00        Length:    178
     Ratio:    2.676         Gaps:       14
Percent Similarity: 41.573   Percent Identity: 30.899

alignment_block:
US-09-664-641-10/rev x B38346

Align seg 1/1  to: B38346  from: 1  to: 223

1775 AGCAGTGTGCTTATCAGACAICTGCTCTGGATAATCCGCAATTGCAAACAC 1726
     |||||||||   ::: :||   ||:||
  77 SerSerCysCysLysProCysCysGlnSerSerCysCys........... 90

1725 ACATCCCAAATAAGAACGCCTTTCTTCTGGAATCTCCACTGCTTGGAATCAHGTC 1676
     ||||||
  91 .........LysProCysCysSerSerGlyCysGlySerSerCysC 103

1675 CAAAAAGCTTCATCGTCCTCGAGGCGATCGTGGTGCGTGCTTGC 1632
     :::||| ||||||||
 103 ysGlnSerSerCysCys.........LysProCysCysGlnSer 115

1631 TGCTGCTGC..........TGCTGTGAAAATGGATGCCGGC...GGCTGCTGGGG 1591
     |||||||  |||   :::||:   |||||||||
 116 SerCysCysLysProCysCysSerSerGlyCysGlySerSerCysCysGl 132

1590 GTAAGGTTGCTGAGACAICTGCGTCGCTGC......................T 1562
     :::||| ||||||||| ||:||
 132 nSerSerCysCysLysProCysCysGlnSerSerCysCysLysProC 149

1561 GCTCCTGCCTGCGAGC......TGCATWTATCTGCTGGGGCTGCGAAGGTGTAAA 1518
     |||||||| |||  :::::::   |||||||
 149 ysCysCysGlnSerSerCysCysLysProCys.........CysSer 161

1517 ACCGGGTGC............................................TGCTG 1504
     |||||||
 162 SerGlyCysGlySerSerCysCysLysProCysCys.........TGCTG 178

1503 CTGC..........TGCTGCTGGCCCTGCCCTGCTGCTGC..........TGCTGTA 1472
     ||||                                          ||||
 178 sCysGlnSerSerCysCysLysProCysCysSerSerCysCysGlnSerSerCysCysL 195

1471 GCAATGCTGTGCTCTCGGAGTCGCACTWTTCACTTCCCTAAACAGCACTGCATTG 1422
     :::|||  |||
 195 ysProCys.........ValCysCysGlnCys 221

1421 GCATTTGTATGT...CCCTGCTGGCTGTGATTCACCTGCTGTTCTWAAATT 1375
     |||||||       ||||       ::|||:|:
 201 SerSerCysCysLysProCysCysCysGlnSerSerCysCysAlaPro.. 216

1374 TTTCGTAGGTGCTGAAAAGAGCTGTGTAAGATATGT 1341
     ||||||
 217 .........ValCysCysGlnCys 221

seq_name: pir2:T14004
```

Right column

```
A:Gene: trfA
A:Introns: 333/3; 364/3; 637/1

alignment_scores:
     Quality:  196.50        Length:    265
     Ratio:    1.456         Gaps:       10
Percent Similarity: 50.943   Percent Identity: 27.170

alignment_block:
US-09-664-641-10 x T14004

Align seg 1/1  to: T14004  from: 1  to: 1390

1207 ATGGCTGCTGGGACAAAACCTCCAAAACTTCTGAAAGATCAGAAATCATACC 1256
     |||||||   ::: :||||
  12 MetAlaAlaGlnGlnGlnSerMetValAlaGlnGlnGlnGlnIleAl 28

1257 TACCTGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATAATGCTG 1306
     :::||| :||::
  28 aGlnGlnGlnSerAlaIleAlaGln........................ 36

1307 ACATTCAGCAGATGAACCGGCCCATCAAAATGTACCACACATATCTTACAGACT 1356
     ||||:||       ::: :|||
  37 ....GlnGlnIleAlaGlnGlnGlnIleAlaGln............... 47

1357 CTTTCAGCACCTACGAAAAATTTAGGACAGCACGGTGAATCACACGACCA 1406
     :||||                   ::::||
  48 ..........................GlnGlnIleAlaGlnGlnGlnGlnGl 56

1407 GGGACATACAAAATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTC 1456
     ::::||  |||||:|     |:||     :::|||
  56 nGlnHis.........GlnGlnIleAlaGlnGlnHisG 65

1457 CAGGACACACAATGCTACGACGCAGCACCAGCCCCAGCAGCAGCAGCAC 1506
     |||    :::|||||    ||:||||
  65 lnGlnHisGlnHisGlnGlnGlnHisGlnGlnGlnHisGlnHisGln 81

1507 CAGCAC.............CCGGTTTTACACCTTCAGCGCAGCAGAT 1541
     |||                     ::: |||||||
  82 GlnHisIleGlnGlnHisGlnProGlnGlnHisGlnProGlnGlnGl 98

1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGCAGAGATCTCTCAGCCAACCTTACC 1591
     :::|||||||||||||||||||||||||
  98 nGlnSerGlnGlnGlnGlnGlnGlnHisGlnGlnGlnHisGlnProGlnGlnHisG 115

1592 CCCAGCAGCCGCCGCCGCCATCCATTT 1614
     ||||||||||||:|||:||
 115 lnGlnGlnProGlnGlnGlnHisGlnProGlnGlnGlnProGln 131

1615 ......TCACGCAGCAGCAGCAGCAGCAGCCACCACCACCGCTCA 1658
     :||||||||||||||||||||||||||||
 132 GlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnProGl 148

1659 GCAGCATCAGCTTTTTGGACATGATCCAGCAGTGGAGAGTTCCAGAAGAAG 1708
     ||||:::||||     |||||||:|||:
 148 nGlnGlnGln.............GlnGlnGlnGlnGlnA 157

1709 GCTTCTTTATTCGGGATGTGTCGTTTGCCAATTGCCGGAT 1743
     |||        :||         |||
 157 spGlnLeuSerGlyValThrGlyLysIleArgAspLeuAsnGluSerIle 173

1744 ..............TATCAGACAGATGTCTGATWAAG...CAACT 1772
     |||:::|||:::|||:
 174 TrpIleHisLeuGlyGlyTyrArgAlaGluSerIleGlyGluAspLysProAl 190

1773 GCTGGCACCTGGAAAAGGATAAATCCAGGACACAAGGCGGCGCACTGTTGACC 1822
     ||||||:::::::|||:::
 190 aLeuAlaSerTyrGlyAlaAsnAlaLeuArg...His...........As 202

1823 CCACCTTCACGAGTCCATCCACCCACCTTCTCTGTGAGAGTCAAGTCAGC 1872
     |||:::::::
 202 roPheSerIleLysAlaLeuThrGlnIleAlaSerLeuPheArgIleLys 218
```

1873 ACCCCCTATCCACCAGGCCAATAACAGAAGAAAGACAAAGAGACGATGTGTTACT 1917
     219 GluGlnThrSerLysAlaAlaGlyThrPheGlnArgIleValThr 233

seq_name: pir2:T03455

alignment_scores:
             Quality: 196.50   Length: 375
             Ratio: 1.062   Gaps: 15
Percent Similarity: 49.333   Percent Identity: 25.067

alignment_block:
US-09-664-641-10:

Align seg 1/1 to: T03455   from: 1 to: 4957

 835 GAAGAGGAAGAGAGGAAGGAAGAGAAGGAGAAGGAGGAGGAG..........GAAGT 872
     :||||||||||||||||||||||||:||||||||||||||||::::::::::||||
3155 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeuLeuLeuGlnVa 3171

 873 AGAAAATGAGGAACAAGATTCTCAGAAATGCAGGGT...AGTACAGATGAGA 919
     :||||||||:|:||:|:||::|||::||::|||||| :::||||||:||||:
3171 lAlaIleGlnGlnGlnGlnGlnGlnProGlyValGlnThrAsnGlnA 3188

 920 AGTCAAGCCCCTGCCAGCTCTCAAGAAGGGTCTCCTTCAGGTGAC......963
     ::|:||| ||||||:|||
3188 laIleGlyProLysProGlnGlyLeuMetProProSerSerHi:GlnGly 3204

 964 .........CACCAGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGG 1004
     :::::::::|||||:|:|||:|||||||||||:|||:||||||||:::
3205 LeuLeuValGlnGlnLeuSerProGln.......................3213

1005 GGAATTAATGTTGATCATTCTTCAGATTCATCACCGGAAAAACAGGAGA 1054
     ||||||::|||:|::||:|::|||||||||||||||||||||||||
3213 3213

1055 GAAATTAAACTGGACCCCGGCCGAAGTCCCACAG.....TTAGCTGCA 1098
     ||| ||||||| :|||:||:
3214 ....ProProGlnGlyProGlnGlyMetLeuGlyPro 3224

1099 GCAAAAACCAGGCTGCCTCAGGGAAAGGAGACCTGGGTTGATTAACTTGTG 1148
     |||:| ||| ||||||||||||||||||||||||||:::
3225 AlaGlnValAlaValLeuGlnGlnGlnHisProGlyAlaLeuGly......3239

1149 TGCCAAATCTCCCACCCGTCCACCGTCTAACACATTTTGCCCCCTGAGGTCCGGG 1198
     ||| ||| |||
3240 ....ProGlnGlyProHis...................ArgG 3246

1199 GTAATTTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAA 1248
     |||||||:::   :::| ||| :::
3246 loValLeuMetThrGlnSerArgValLeuSerSerProGln.........3259

1249 ATGATAGCTTACCTGGAGTCCACCTGTACCGGACACACTCAGGAATATTACTAA 1298
     ::||::|| ||| ::||| ::|||
3260 .....LeuAlaGlnGlyGlnGlyLeuMetGlyHisArgLeuValThrAl 3275

1299 TAATGCTGACATTCAGCAGATGAACCGGCCCATCAAAATGTACACATATCT 1348
     ::: |||||||:||| ::|||
3275 aGlnGlnThrSerLysAlaAlaGlyThrPheGlnArgIleValThr 3233

     ::: |||||||||::::||
3275 aGlnGlnThrSerLysAlaAlaGl...........HisGlnG 3286

1349 TACAGACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCAC 1398
     ||| |||||||:::::| :::||:||:|||
3286 lnGlnGlySerMetAlaGlyLeuSerHisLeuGlnHisLeuSerLeu...Met 3301

1399 AGCCACAGCAGGACATACAAATCCAATGCAGGTGCTGTTTACCAAGTGAA 1448
     |||:::|||||||:|||||||||| |||:::|||||||::::::|||
3302 SerHisSerGlyGlnProLysLeuSerAlaGlnProMetGlySerLeuGl 3318

1449 AGTGCACTCCAGAGACAGACATCCTACAGCAGCACCAGCACCCCAGCAGC 1498
     :||||||||||||||||||:|||||||||||||||||||||:|||||
3318 nGlnLeuGlnGlnLeuGlnGlnLeuGlnLeuGlnGlnLeuGlnGlnLeuGlnG 3335

1499 AGCAGCAGCAG......CACCCGGTTTTACACCTTCAGCCCCAGCAGATA 1542
     |||||||||| :::||||||||||||||||||||:||||
3335 lnGlnLeuGlnGlnLeuGlnLeuGlnLeuGlnLeuGlnLeuGlnGlnLeu 351

1543 ATGCAGCCTCCAGCAGCACCAGCAGCAGCAGATCTCTCAGCAAACCTTACCC 1592
     |||||||:::|||||||||||||||||||||||:||:|||||||||||||
3335 GlnGlnLeuGlnGlnLeuGlnGlnLeuGlnGlnLeuGlnGlnLeuGlnGlnGlnIle 3368

3352 GlnGlnLeuGlnGlnLeuGlnLeuGlnGlnLeuGlnGlnLeuGlnGlnIle 3368

1593 CCAGCAGCCGCCCGCGATCCCATTTTCACAGCAGCAGCAGCAACAGCAGC 1642
     ::: :::||||||||||||:::|||||||||||||||||||||||||||
3368 uGlnLeuGlnGlnLeuGlnGlnLeuGlnLeuGlnGlnLeuGlnGlnLeuGlnGlnLeu 3385

1643 CACCACCATCGCCTTCACCAGCACAGCATCCAGCTTTTTGGACAATGATCCAGCAGTG 1692
     :::||||||||||||||||||||||||||||||||||:||:|||:||||||||
3385 uGlnLeuGlnGlnLeuGlnGlnLeuGlnMetGly.................3396

3385 InPheGlnGlnGlnGlnGlnLeuGlnMetGly.................3396

1693 GAGATTCCAGAAGAGGCTTCTTATTGGGATGTGTGTTTGCAATTGCGGA 1742
     ::::::||||||||:|||||||||||||||||||||||||||
3397 LeuLeuAsnGlnSerArgThrLeuLeuSer....................3406

1743 TTTATCCAGACGACAGTGTCTGATTAAGCAA................. 1770
     |||:||||||||||||||||||:::
3406 ....ProGlnGlnGlnGlnGluValAlaAlaLeuGlyProGlyMetP 3422

1771 CTGCTGGCCACCTGGAAAAGGATAATCCAGGCACATGGCGGCCACTGTTGA 1820
     ||||| :|| :||||||||||||||||||||||||||||||||||
3422 roAlaLysProLeuGluGlnHisPheSerSerProGlyAlaAlaLeuGlyProThr 3438

1821 CCCCACCTTCACCAGTCGGATCCACG 1845
     ||||:::||||:|::|||:|||||||
3439 LeuLeuLeuThrGlyLys.........GluGlnAsnThrValAs 3450

3450 pProAlaValSerSerGluAlaThr 3458

OM of: US-09-664-641-10 to: Issued_Patents_AA:* out_format : pfs

Date: Jun 4, 2002 5:33 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL-frame1_n2p.model -DEV-xlp
-O-/cgn2_1/USPTO_spool/US09664641/runat_04062002_110932_9169/app_query.fasta_1.3597
-DB-Issued_Patents_AA -QFMT-fastan -SUFFIX-rai -GAPOP-12.000
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
-XGAPOP-4.500 -QGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
-PGAPOP-6.000 -PGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-blosum62
-TRANS-human40.cdi -LIST-45 -DOCALIGN-200 -THR_SCORE-pct
-THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs
-NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000000
-USER-US09664641_@CGN1_1.44 -NCPU-6 -ICPU-3 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXY -WAIT -THREADS-1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: Issued_Patents_AA:*
Database length: 24425594
Database sequences: 231628
Search time (sec): 55.080000

score_list:

Sequence	Std Orig	%Score	EScore	Len	! Documentation
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-977-767-3	236.50	356.91	1.6e-13	1345	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-630-915A-37	236.00	355.54	1.8e-13	1400	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-508-761B-22	224.50	335.72	2.3e-12	1417	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-728-323A-2	214.00	320.29	2.0e-11	1162	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-6	197.00	306.02	4.0e-10	362	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-906-349A-6	196.00	294.23	8.2e-10	801	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-415-751-35	194.00	300.89	7.7e-10	362	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-814-964-13	184.00	278.67	9.0e-09	542	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-258-442-13	184.00	278.67	9.0e-09	542	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-328-308-8	184.00	278.67	9.0e-09	542	!
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US92-11107-13	184.00	278.67	9.0e-09	542	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-990-823-112	182.50	293.02	5.4e-09	143	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-916-352-2	174.00	253.74	1.2e-07	1004	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-508-761B-22	173.00	279.00	3.9e-08	120	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-625-188-20	173.00	256.09	1.2e-07	729	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-853-310-2	173.00	253.37	1.4e-07	903	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-267-003B-9	169.00	247.82	3.1e-07	816	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-17	169.00	247.82	3.1e-07	816	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-910-925-3	162.50	237.89	1.2e-06	743	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-2	159.50	241.65	1.5e-06	369	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-991-300-2	159.50	241.65	1.5e-06	369	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-918-914-4	158.00	229.45	3.4e-06	788	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-09-115-446-2	153.00	221.85	9.7e-06	731	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-28	152.50	225.49	8.7e-06	513	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-29	152.50	225.08	8.8e-06	530	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-30	152.50	224.56	9.1e-06	552	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-31	152.50	223.74	9.4e-06	589	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-961-083-56	152.50	219.92	1.1e-05	796	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-246-982A-56	152.50	202.47	2.7e-05	3144	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-453-265-6	152.50	202.47	2.7e-05	3144	!
/cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-457-273B-42	152.50	202.47	2.7e-05	3144	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-556-419-21	152.50	202.47	2.7e-05	3144	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-041-886-15	152.50	202.47	2.7e-05	3144	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-227-536-2	151.50	204.12	2.9e-05	2414	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-910-925-4	151.00	204.12	2.9e-05	2414	!
/cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-04682-2	151.00	218.93	1.5e-05	703	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-194-468-2	150.50	202.26	3.6e-05	2441	!
/cgn2_6/ptodata/2/iaa/6A_COMB.pep:US-08-961-739-2	150.50	202.26	3.6e-05	2441	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-185-432-2	149.50	215.76	2.1e-05	737	!
/cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-185-432-4	149.50	215.76	2.1e-05	737	!
/cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-910-925-1	149.00	215.26	2.3e-05	717	!

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-977-767-3

seq_documentation_block:
; Sequence 3, Application US/08977767
; Patent No. 5972684
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Yue, Henry
; APPLICANT: Greenwald, Sara
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/977,767
; FILING DATE: Herewith
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0423 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1532042
; US-08-977-767-3

alignment_scores:
Quality: 236.50   Length: 194
Ratio: 2.489   Gaps: 13
Percent Similarity: 48.969   Percent Identity: 32.990

alignment_block:
US-09-664-641-10/rev x US-08-977-767-3 ..

Align seq 1/1 to: US-08-977-767-3   from: 1 to: 1345

      1778 GCCAGCAGTTGCTTATCAGACATCTGCTCTGGATAATCCGCAATTGCAAA 1729
           ||||:||||||:||| .:|| ||||| |::|| |:||||
      1057 AlaAlaThrCysThrCysGlyGlyThrCysCysAlaGi 1073
           :|||

      1728 CACACATCCCAATAAGAAGCCTCTCTGGAATCCCACTGCTGGATCAT 1679
           :||| :|||:::||| :: ||||||| |

```
seq_documentation_block:
Sequence 37, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
  APPLICANT: SPARKS, Andrew B.
  APPLICANT: HOFFMAN, No. 6309820h
  APPLICANT: KAY, Brian K.
  APPLICANT: FOWLKES, Dana M.
  APPLICANT: McCONNELL, Stephen J.
  TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
  TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
  TITLE OF INVENTION: USING SAME
  NUMBER OF SEQUENCES: 227
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Pennie & Edmonds LLP
    STREET: 1155 Avenue of the Americas
    CITY: New York
    STATE: New York
    COUNTRY: USA
    ZIP: 10036-2711
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/630,915A
    FILING DATE: 03-APR-1996
    CLASSIFICATION: 536
```

```
ATTORNEY/AGENT INFORMATION:
  NAME: Misrock, S. Leslie
  REGISTRATION NUMBER: 18,872
  REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (212) 790-9090
  TELEFAX: (212) 869-8864/9741
  TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
  LENGTH: 1400 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA
US-08-630-915A-37
```

```
alignment_scores:
  Quality:     236.00          Length:      228
  Ratio:         2.408         Gaps:         14
Percent Similarity:  42.982   Percent Identity:  28.947

alignment_block:
US-09-664-641-10/rev x US-08-630-915A-37

Align seq 1/1  to: US-08-630-915A-37 from: 1  to: 1400
```

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMH.pep:US-08-900-230-3

```
486  AlaCysThrGlyGlyCysGlyCysGlyAlaCysThrThrCysAlaGlyCy 502
     |||:|::|||                               ::
1419 ATTTGATGTCCCCTGCTGCTGCTGTGATTCACCTGCTGCTTCTAAATTTTCG 1370

502  sThrCysCysAlaCysGlyThrGlyGlyCysGlyCysCysAla........ 515
     |||||||::
1369 TAGGTCCTGAAAAGACTCTCTAAGATATGTGCT 1338
     |||  |||||
516  ............CysCysAlaGlyCysAla 521
```

; seq_documentation_block:
; Sequence 3, Application US/08900230
; Patent No. 6329197
; GENERAL INFORMATION:
; APPLICANT: Baird, Jonathan A.
; TITLE OF INVENTION: DNA ENCODING GALANN GALR3 RECEPTORS AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,230
; FILING DATE: 23-JUL-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52241-C/JPW/ADM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1417 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE:
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

alignment_scores:
              Quality: 224.50   Length: 150
                Ratio: 3.454   Gaps: 10
   Percent Similarity: 43.333   Percent Identity: 36.667

alignment_block:
US-09-664-641-10/rev x US-08-900-230-3

Align seq 1/1 to: US-08-900-230-3 from: 1 to: 1417

```
1685 GGATCATGTCCAAAAAGCTGCATCCTGCTGCTGAGGCGATGGTGGGTGGCTGC.. 1638
     ||||||   |||||||  ||||||   ||||||   ||||||
1064 GlyCysCysAlaCysThrThrCysCys...GlyCysGlyCysGlyCysGl 1079
     ||||||         |||||||
1637 ............TGTTGCTGCTGCTGC.............T 1622
     |||||||
1079 yCysThrThrCysGlyCysCysGlyCysCysCysThrGlyThrGlyGlyC 1096
```

---

seq_name: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:US-08-728-323A-2

; seq_documentation_block:
; Sequence 2, Application US/0872832A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1162 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
1621 GCCTGTGAAAATGGAATGCGGCGCGGCGTGCTGCGGGG..............TAAGGT 1584
     |||||  ::||||||||||||||||||  |||    |||
1096 ysCCysGlyThrGlyCysGlyGlyThrGlyCysCysGlyGlyAlaCysGly 1112

1583 TGCTGAGAGATCTGCCTGCTGCTGCTCGC........TGCTGCTGCGAGCTTGCAT 1543
     |||:::|||                              :::|:|||
1113 CysCysGlyCysCysGlyAlaCysGlyThrGlyCysCysGlyCysCysGlyCy 1129

1542 TATCTGCTGGGGCTGAAGGTGTGAAAACGGGGTGC.............. 1509
     |||      ||||:|||||
1129 sGlyCysGly.......CysCysThrThrGlyCysGlyThrCysGlyCysG 1144

1508 ...........TGCTGCTGCTGCC...TGCTGGGCCTGCGTGC........ 1482
     ||||||||||                ||| :::|||||||
1144 lyThrCysCysGlyGlyGlyCysGlyCysCysGlyCysGlyThrCysCysThrCys 1160

1481 ........TGCTGCTGTAGCATGTGTGTCTGTCTCTGGAGTGCACTTCACTTG 1441
     |||||||||   |||||||||  ||| |||||
1161 GlyGlyGlyCysGlyCysCysGlyAlaCysGlyCysGlyThrCysGlyGlyCy 1177

1440 GCTAAACAGCACTGCATTGGGCATTGTGATGTCCCTGTGCTGGGCTGATTCA 1391
     |||       |||  |||||||||
1177 sCysGlyCysGlyLyAlaGlyAlaCysGlyCysGlyCysGlyCysGly...... 1191

1390 CCTGCCTGTTCTAAAATTTTTCGTTAGGGTGCCGAAAAGAGCTCTGTAAAGAGATGATGT 1341
     ||||:::||             |||    |||
1191 LyCysCysThrAlaGlyLyCysGlyLyGlyAlaGlyLyCysGlyCysThrGlyCys 1207
```

US-08-728-323A-2

alignment_scores:

```
          Quality:   214.00        Length:    317
          Ratio:     1.417         Gaps:       11
Percent Similarity: 47.634   Percent Identity: 24.606
```

alignment_block:
US-09-664-641-10 x US-08-728-323A-2

Align seg 1/1 to: US-08-728-323A-2 from: 1 to: 1162

```
835 GAAGAGGAAGAGGAGGAGGAAGAGGAAGAGGAAGAGGAAGAGCTACGAAAATGAGGA 884
    ||||||||||||::::|||||||||||||||||||||||||:|||:::::||
390 GluGluGluAspGluGluGluAspGluGluGluAspGluGluGluAspGluAspAs 406

885 ACAAGATTCTCAGAATCGAGGCT....AGTACAGATCAGAAGCTCAAGCCCTG 931
    :::::||||||||||::::    :::|||||||||||:::
406 pAspAspAspGluAspAsnGluAspGluCluAspGluCluAspGluCluAspAs 423

932 CCAGCTCTCAAGAAGGGTTCCTTCAGTGACCACGCAGTTTTCACCTAAA 981
    ::::::::||||:                :::|||::||:
423 ysGluAspGluGlyGlyAspGlyAsnLysThrLeuSerIleGln 439

982 TCCAACACTGAAAAATCTAAAGGGGAATTAATTTGTTGATCATTCTTCAGA 1031
    |||:|||                    ::::|||||::||
440 SerSerGlnHisGlnGln.......GluProGlnGlnProGlnGl 454

1032 TTCATCACCGAGAAAAACAGAGACAGAAATTTAAACTGGACCCCGGCGAAG 1081
    :::::::::|||||||||::::|||||||:|||
454 nGlnGlnProGlnGlnGlu.......ProLeuGlnG 465

1082 TCCCACAGTTAGCTGCAGCAAAAACGCAGGCTGCCCTCCAGGGAAAGGAGCCT 1131
    ||||||                    :::|||||||||
465 luProGlnGlnGlnluProGlnGlnGlnGlnGlnGluPro 481

1132 GGGTTGATTAACTTGTTGTGCCAATGTCCCACCGTGTCCCAGGTAACATTTT 1181
    :::::::|||                ::||||||
481 luProGlnGlnGluluProGlnGlnGlnGlnGlnGluPro 481

1182 CCCCCCTGAGCTCCGGGGTAATTTAATGGCTGGTCGACAAAAACCTCCAA 1231
    |||||||||                    |||||:
481 .................................ProGlnGlnGlnGlu 481

1232 GTTCTGAAAGATGAGAAATGATAAGCTACCTGGAGTCCCACCTGTACCGGACA 1281
    ||||||                    |||||||
482 .................................LeuGln 484

1282 CTGAGGAATATTACTAATAATGCTGACATTCAGCAGATCAACCGGCCATC 1331
    |||                        :::|||
484 luProGlnGlnGlu.......................... 489

1332 AAATCTACCACATATCTTACAGACTCTTTCAGCCACCTACGAAAAATTTAG 1381
    |||||                    ||||||:
490 .................................ProGlnGlnGlnGlu 495

1382 AACAGCAGGTGAATCACAGCCAGCAGGGACATACAAAATGCCAATGCAGTG 1431
    |||||                    :::|||
496 .................................LeuGlnProGlnGlnProGlnGlnP 504

1432 CTGTTTAGCCAAGTGAAAAGTGACTCCAGACACACACATCTACAGCAGCA 1481
    |||||||                  ||||||||
504 roGlnGlnGluluProGlnGlnGln.................. 512

1482 GCACCAGCCCCAGCAGCCAGCAGCCAGCACCTTCAGTTTTACACTTCAGC 1531
    |:::                      |||||:
513 .................................GluProGlnGlnGluluProGlnGl 522

1532 CCCAGCAGAATAATGCAGCTCCAGCAGAGCAGCAGCAGGCAGCACAGATCTCTCAG 1581
    |||||||                  |||:
522 nGlnGlnProGlnGlnGlnProGlnGlnC n.......Glu 534

534 roGlnGlnGluluProGlnGlnGluluProGlnGln 550
```

us-09-664-641-10.rai

seq_documentation_block:
Sequence 6, Application US/08415751

; Patent No. 5643772

seq_name: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:US-08-415-751-6

```
1582 CAACCTTACCCCCAGCAG......CCGCGCGCCATCCATTTTCACAGCAGCA 1625
     |||             :::    :: :::|||||:::
551 GlnArgGluProGlnGlnArgGluProGlnGlnArgGluProGlnGlnAr 567

1626 GCAGCAGCAACAGCAGCCACAGCCACACCA......TCGCCTCACCAGCATCAGC 1669
     ::::|||||:::      |||    :::|||||:::
567 gGluProGlnGlnArgGluProGlnGlnArgGluProGlnGlnArgGluProGlnGlnArgGluP 584

1670 TTTTTGGACATGATCCAGCAGTGGAGAGATTCCAGAAGAGGAGAGGCTTCTTATTG 1719
     :::::|||||||||:::        ::::|||||:::
584 roGlnGlnArgGluProGlnGlnArgGluProGlnGln.......... 596

1720 GGATGTGTGTTTTGTTCCAATTGCCGGATTATCCAGACAGAGTGTCTGATAAGCA 1769
     ||||:::::|||:::        ::::|||:::
596 .................GlnAspGluGlnGlnAspGluGlnGl 605

1770 A 1770

605 n 605
```

```
;   ORGANISM:  Cryptosporidium parvum
;   FEATURE:
;      NAME/KEY:   Positions coded by nonsense codons are
;      NAME/KEY:   identified as Xaa.
US-08-415-751-6

alignment_scores:
US-09-664-641-10/rev x US-08-415-751-6
      Quality:   197.00     Length:       217
      Ratio:       2.189    Gaps:          16
Percent Similarity: 41.475  Percent Identity: 26.728

Align seq 1/1  to: US-08-415-751-6  from: 1  to: 362  .

1896 TCTTATTGCCTTGTCCATACGGCCCTGTGACTTCACTCACAGAGAAGGT 1847
     ||||||       |||::::|||:     :::             ::||
123  SerTyrLeuSerCysLeuArgThrThrpAsnTrpPheIleIle...ArgVa 138

1846 GCGTGCAATCGACTCGTGAAGGTGGGGTCAACAGTGCCGCCATGTGCCTGG 1797
     :::|||  ||||||      ||||:|||:::|||::||:
138  |********ThrThrVal IleVal IleAsnHisSerCys............ 151

1796 ATTATCCTTTTCCAGGTGGCCACCACTTGCTTATCAGACATCTGCTC... 1750
     ::::|||||                   |||||||||||:::|::
152  .........TGCATA           ValLeuIleGlnHisLeuPheAsn 159

1749 .........TGGATA 1736
     .||       |||
160  TrpAspTrpCysAsnpheAsnValThrAspIleArg***TrpCysArgCy 176

1735 TTGCAAACACAATCCCAATAAGAAGCCTTCTTCTGGAATCCACTGCT 1686
     |||  |||    ||||||||||||||                ::|||
176  sCysCysHis...........                    PheValSerLeu.GlySer 185

1685 GGATCATGTCCAAAAAGCTGAATCGTGCTCAGCCGATGGTGGTGCCTGTG 1636
     ::||||||||||                         ||||||||||||||
186  SerSerCysArgTrpHisCysCysCys............       CysCy 196

1635 TTGCTCCTGCC...TGCTGCTGTGAAAATGGATGCCCGCCGCTGCTGGGGGT 1589
     ||||||||||      |||||||||||||||||||||||||||||||
196  sCysCysCysPheCysCysCys............       CysCysAsnTyrThr 208

1588 AAGGTTGCTGAGACAGATCTGCTGCCTCGTCGTGGAGGCTGCAATATC 1539
     :::|||||||||||||||||||||||||||
209  ...........Leu***CysCysCysCysGlyTyrTyrTrpTrpLeu 220

1538 TCCTGGGGGCTTGAACCTGTAAAACCGGGTCGCTGCTGGTTGCTGCTGGGC 1489
     ::::||||||       ||||||||||||||
221  ThrTrp...........       AsnLeuCysCysCysCysTrpPh 232

1488 CTGCTCCTGCTGCTGTGTAGCATGTGTGTCTCTGGAGTCACTTTCACTTGC 1439
     ||||||||         ||:::|||:
232  eLeuSerCysCysAsn...........       .SpTrp 241

1438 TAAACAGCACTGCATTGGCATTTGTATGCCCTGCGTCGTGATTCACC 1389
     ::::|||        ::            |||
241  LeLysSer...........       CysCysGlyCys 247

1388 TCCTGTTGTCTAAATTTTTCGTTAGGTGCTGAAAGAGTCTGTAAGATATGTGC 1339
     |||||||       :::          ::::
248  CysCysLeuArg***Trp********** 260

1338 TACATTTGATTGACGCGGTTCATCTCCTGAATGTCAGCCATTATTAGTAATA 1290
     ||:::      |||:::
260  uAspPheGln...........       LeuValCysle 270
```

```
;   Patent No. 5434064
;   GENERAL INFORMATION:
;   APPLICANT: Schlessinger, Joseph
;   APPLICANT: Skolnik, Edward Y.
;   APPLICANT: Margolis, Benjamin L.
;   TITLE OF INVENTION:  A NOVEL EXPRESSION-CLONING METHOD FOR
;   TITLE OF INVENTION:  IDENTIFYING TARGET PROTEINS
;   TITLE OF INVENTION:  TARGET PROTEINS
;   TITLE OF INVENTION:  FOR EUKARYOTIC TYROSINE KINASE
;   NUMBER OF SEQUENCES:  16
;   CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Browdy and Neimark
;      STREET:  419 Seventh Street, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;   COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/906,349A
;      FILING DATE:  30-JUN-1992
;      CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/643,237
;      FILING DATE:  18-JAN-1991
;   TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202-628-5197
;      TELEFAX:  202-737-3528
;   INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;      LENGTH:  801 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-07-906-349A-6

alignment_block:
US-09-664-641-10/rev x US-07-906-349A-6

alignment_scores:
      Quality:   196.00     Length:       420
      Ratio:       1.273    Gaps:          22
Percent Similarity: 36.667  Percent Identity: 22.381

Align seq 1/1  to: US-07-906-349A-6  from: 1  to: 801  .

2257 GAATCTGCCTCAGTGCCTCAAAGTTCCA.................. 2228
     |||||||||||||||||||||           :::
87   GluSerGluSerAlaproGlyAspPheSerLeuSerValLysPheGlyAs 103

2227 ..................GAAGAATGTGCCCAAGCCACTCGGGGTTGACACA 2194
                      ::: |||::||||||:::||||||||||||||
103  nAspValGlnHisPheLysPheLysValLeuArgAspGlyValaGlyLys... 117

2193 GGGTAFCCTCCACTCTTTGGCTT................. 2150
     |||||                   |||:::|||::||||:
118  ..TyrPheLeuTrpValVallysPheAsnSerLeuAsnGluLeu 131

2149 GTTCTTTTACAGATGAGGAGCTGTGTTGCTCGCGCAATAGATAACCCGTAT 2102
     ||:::      :::||||||||||||||||||||||
132  ValAspThrTyrHisArgSerThrSerValSerArgAsnGlnIleIlePheLe 148

2101 ....................ATTTGGCACCTGCCAAATAAGCCATTAATTTTA......... 2069
                      :::|||:|||:||||||:||||         :::
148  uArgAspPleGluGlnValProGlnGlnProThrTyrValGlnAlaLeuP 165

2068 ...................................................GGTCAT 2063
                                                       |||
```

```
 165 heAspPheAspProGlnGluAspGlyIleArg*GlyPheArgGlyAlaAsp  181
2062 CTCTGCTCACTATCAACAAAATCCAGTCACAGAAATAATATCCTGTCAACAT 2013
          :::                        :::
 182 PheIleHis.ValMetAspAsnSerAspProAsn..................  192
2012 GGCTTTTCCTCCTGGTGGGAAGGC...............CACTGGAAGTGAAGGGCTCG 1969
     |||||||:::|||                        ||| :::::
 193 ..........TrpPheLysGlyIleAlaCysHisGlnThrGlyMetPhe  205
1968 GTCCGCGGGTACCATTTTTTTTTTTTTTTTTAAGACTGTGTTTAACCAGTGTG 1919
     |||| ||||||
 206 ProArgAsnTyr.......................ValThrProValAs  214
1918 CACTAAC...............ACATCTCTTCTTTCTTTCTTATT 1890
     |||                   :::
 214 narAsnValCysAlaAlaIaAlaIaCIyAlaAlaThrThrAlaAlaAlaC  231
1889 GGCTGTGCATACCGCGCTGCTGACTTGACCTCACAGAGAAGGTGCGTGCA 1840
     || ||| |||    |||    :::
 231 ysCysCysCysAlaAlaCysAlaAlaIaGlyCysThr..........GlyCys  242
1839 TCCACTCGTCGAAGGTGGGCGGTCAACACGTGCCCGCATGTGCCTGGATTATCC 1790
     :::||||  |||:::   |||
 243 CysThrCysThrGlyAlaCysAlaGlyCysAlaGlyCysCysThrGlyTh  259
1789 TTTTTCCAGGT.........GGCCACCAGTTGCTTATCACGACATCTGCTGC 1749
     |||                 ||  :::|||
 259 rGlyAlaGlyGlyAlaGlyAlaGly.ThrGlyCysAlaGlyAlaAlaCysAla  275
1748 GGATAATCCCGCAATTGCAAACACACATCCCAATAAGAAGCCTTC 1705
     ||| |||   :::||| |||:::||||||
 276 CysCysGlyThrThrThrThrCysThrThrThrAlaIaAlaaThrThrThrGlyI  292
1704 TTCTGGAATCCACTCGACTGCTCGGATCATGTCCAAAAAGCTGATGCTGCTGAG 1655
     :::|||   :::    |||    |||
 292 yGlyThrGlyThrThrThrCysGlyGlyAlaCysGlyGlyAlaAlaCysThrG  309
1654 CCGAT.........GGTGGTGGCTTGCCTGGTTGCTGCTGCTGC... 1620
     ||                   |||||||:::|||    |||||||||
 309 lythrCysGlyThrCysGlyAlaCysGlyAlaAlaCysThrThrCysCysThr  325
1619 ...............TGTGAAAATGGAGGCGCGGCGGC 1597
                    |||:::|||
 326 CysAlaCysGlyThrThrThrThrThrAlaGlyIyGlyThrGlyIyGlyGl  342
1596 CTGGGGGTAAGGTTGCCTGACAGATCTGCCTGCTGC............. 1562
     :::::|||        :::|||||||||
 342 sGlyGlyGlyThrCys.....AlaCysSysCysGlyThrThrGlyIyAlac  357
1561 GCTCCTGCTGCTGGAGCTGCCATTATCTGC................ 1536
     |||||||                  |||
 357 ysCysCysThrCysThrGlyGlyIyThrThrGlyIyThrThrThrGlyIyThrThrGly  373
1536 .................................................. 1536
 374 GlyAlaAlaIaAlaCysThrThrThrAlaIaGlyIyGlyGlyGlyIyGlyGl  390
1535 ................TGGGGCTGAAGGTTGTAAAACCCGGGTGCTGC...... 1506
                     |||||||||||||
 390 yAlaGlyGlyIyGlyGlyIyGlyCysGlyAlaCysGysGysGlyAlaAlag  407
1506 .................................................. 1506
 407 lyThrGlyIyGlyAlaIaaCysThrGlyGlyIyGlyAlaGlyIyAlaIaaGly  423
1506 .................................................. 1506
 424 ThrGlyIaAlaAlaAlaCysCysGlysAlaAlaCysGysCysGysCysGysGysGysGysCysGysThrGlyIyAlaAlaAl  440
1505 ...............TCGCTGCTTGGCTGCTGCTCGGGCCTCTGCTGGCTGCGTAGCACATGTGTTCTT 1460
          ||||||||||||        |||||||||||         ||| :
 440 aThrCysCysCysCysGysAlaCysGysGysCysGysGysGysGysThrCysCysCysCysGysG  457
```

seq_name: /cgn2_6/ptodata/2/iaz/5A_COMB.pep:US-08-415-751-35

```
1459 CTGGAGTCACTTCACTCGGCTAAAACAGCACACTGGATTGGCATTTGTATGT 1410
     :::|||||::|||    :::||||||:::||||    |||    :::
 457 lyGlyThrThr...GlyGlyAlaThrThrThrAlaAlaAlaAlaAlaThr  472

1409 CCCTGC 1404
      |||
 473 GlyCys 474
```

alignment_scores:
      Quality: 194.00   Length: 232
        Ratio: 2.337   Gaps: 13
Percent Similarity: 35.776   Percent Identity: 22.845

alignment_block:
US-09-664-641-10/rev x US-08-415-751-35 .

Align seq 1/1 to: US-08-415-751-35   from: 1 to: 362

1664 TGCTGCTGAGGGCATGGTCGTGGCTGCTGT................................
      |||||||
115   CysCys........................GlyCysCysSerLysIleTrpAspAsnCy 126

1634   TGCTGCTGCTGCTGCTGCTGTGTGAAAATGGANTGCG 1604
       |||||||||||||
126   sCysCysSerLeuPheThrTrpCysCysCysCysCysCysLeuAsnTrpThr. 142

1603 GCGGCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGCC 1554
       |||||||| |||
143   ....LeuTrpTrpTyrGlyCys...........CysCysCys***Leu 154

1553 TGG...AGCTGCATTATCTGCTGCTGGGCGGCCTGAAGGTTGTAAAACGGGTGCTG 1507
       ||| ||||| :::|||
185   CysCysCysArgCysTrp***ArgIleAspThrValAsnPheAsnCysTr 201

1383 TTCTAAATTTTTTGGTAGGTGCTGCTGAAAAGAGTTCGT............. 1350
       :::|||
201   pAspTrpAsnPheLeuHisGlnIleLeuValCysSerHisAsnIleValI 218

1349   AAG 1347
                                  :::
218   leValThrThr*********ValArgIleIlePheTrpAsnTrpThrArg 234

1346 ATATGTGCTACATTGATGGCGCGTTCATCTGCTGAATGTCAGATTATT 1297
       :::||| |||::|||
235   LeuCysSerLeuTyrSer............IleCysAlaIle***GluTrpTh 248

1296 AGTAATAATTCCTCAGTGTC.............................. 1278
      :::
248   r****Val***LeuLeuValProLysGlyIleThrGlyIleThrProArgG 265

1278   1278
265   lyAsnSerGlyPheAspProLeuLysThrCy.GlyThrValLeu*** 281
                        |||
1277 .CGTACAGTGGACTCCAAGTAGCTAT 1252
      |||:||| |||:||:
282   IleAspProLysAspThrAspTyrArgAlaIleAlaLeuPheLysValThrSe 298

1251 CATTTCTGATCTCTTCAGAACTTTGGAGGTTTGTCCAGAGCCATT 1206
       |||::|||
298   rValSerSerValValThr***TrpAspIleCysGluSerThrVal 313

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-07-814-964-13

seq_documentation_block:
; Sequence 13, Application US/07814964
; Patent No. 5355047
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven

; APPLICANT: Aellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Grauahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 542 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: fractional yeast SSR (fYSSRP) (predicted)
US-07-814-964-13

alignment_block:
US-09-664-641-10 x US-07-814-964-13

Align seq 1/1 to: US-07-814-964-13   from: 1 to: 542

alignment_scores:
                 Quality: 184.00   Length: 509
Percent Similarity: Ratio: 0.786   Gaps: 23
                                   Percent Identity: 20.236

832   TATCAAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGTAGAAAATGA 881
      :|:||:   :::::::||:::::::||::::::: :::::
4   PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 20

882   GGAACAAGAgTTCTAGAATGAGGGTAGTACAGATGAGATGACGAAGTCAAGCCCCTG 931
      ::||:||||:::::   :|| :::::::::
20   nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln......... 32

932   CCAGCTCTCAAGAAGAGGGTCTCCTTCAGGTGACCAGCAG............. 969
      |||||||:: |||||
33   ....SerProGlnGlnGlnAsnValTyrPhe 43

970   ...TTTTCACCTAAATCCACACTGAAAAAATCTAAAGGGAGAATTAATGTT 1016
      :: |||:||| |||:||::::::::::
44   ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGlnGlG 60

1017 TGATGATTCTTCAGATTCATCACCGGAAAAAACAGGAGAGAAAATTAAACT 1066
      ::::::::|| :::::::::::::::::::||::||::||

```
  60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnValAsnV 77

1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGCAGGCTGCCT 1116
       :::         :::||||||      :::      ::::::
  77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

1117 CAGGGAAAGGAGCCTGGGTTGATTAAC....................... 1143
     :::|||::::::         ||||||
  92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheTy 108

1144 .......TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCC 1186
            :::   :::::::::               |||         |
 108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.........P 122

1187 CTGAGGTCCGGGGT.................................AAT 1203
     ||:::|||   |||                                 |||
 122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138

1204 TTAATGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
        :::::::::|||:::|||      ||||||:::   :::::::::
 139 AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl 155

1254 AGCTACC...TGGAGTCCAGCTGTACGGACACTGAGGAATATTACTAATA 1300
      |||:::      :::||||||:::   |||            :::|||
 155 aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA 172

1301 ATGCT.....................GACATTCAGCAGATGAACCGG... 1326
       |||                     ::::::   |||::::::|||
 172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

1327 ...CCATCAAATGTAGCACATATCTTACAGACTCTTTCAGCACCTACGAA 1373
        |||:::|||:::   :::   |||::::::|||     |||:::::
 189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

1374 A................................................. 1374
     :
 204 rAlaGlyAsnAlaAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

1375 ........AATTTAGAACAGCAGGTGAATCACAGCCAGCAGGGACATACA 1416
             ||||||:::   |||::::::|||   ||||||
 221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln......... 234

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466

 234 .................................................. 234

1467 CATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCAGCAGCAGCACCCGG 1516
      ::::::|||||||||   |||   |||||||||||||||
 235 .GlnMetGlnGlnGlnLeuGlnLeuGlnGlnGlnGlnGln.......... 247

1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG 1566
             ||||||   ||||||            |||   |||||||||
 248 ........LeuGlnGlnGlnGln............GlnLeuGlnGlnGln 257

1567 CAGCAGATCTCTCAGCAACCTTACCCCCAGCAGCCGCCGCATCCATTTTC 1616
     :::|||::::::||||||         ||||||      |||   :::::
 258 HisGlnLeuGlnGlnGlnGlnGlnLeuGlnGlnGlnHisHisHisLeuGl 274

1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA...................... 1644
     :|||||||||||||||||||||:::|||
 274 nGlnGlnGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291

1645 ............................................CCACCA 1650
                                                 |||
 291 hrGlnSerArgIleGluArgArgLysGlnLeuLysLysGlnGlyProLys 307

1651 TCGCCTCAG..............................CAGCATCAGCT 1670
        |||:::                              :::::::::||
 308 ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

1671 TTTTGGACATGATCCAGCAGTGGAGATTCCAGAAGAAGGCTTCTTATTGG 1720
     |:::   :::   |||   :::::::::||||||   :::   |||
 324 uLeuGlnGlnPheProGluAlaLysValProGluLeuSerLysLeu.... 339
```

```
1721 GATGTGTGTTTGCAATTGCGGATTATCCAGAGCAGATGTCTGAT...... 1764
                |||   |||   :::   |||      :::|||
 340 ...........AlaSerAlaArgTrpLysGluLeuThrAspAspGlnLys 352

1765 ............AAGCAACTGCTGGCCACCTGGAAAAGG.........AT 1793
                 :::::::::   ::::::|||::::::         ::
 353 LysProPheTyrGluGluPheArgThrAsnTrpGluLysTyrArgValVa 369

1794 AATCCAGGCACATGGCGGCACTGTTGACCCC................... 1824
     :   :::|||:::      |||:::   |||
 369 lArgAspAlaTyrGluLysThrLeuProProLysArgProSerGlyProP 386

1825 .....ACCTTCACGAGTCGATGCACGCACCTTCTCTGTGAGAGT...... 1863
             ||||||::::::            :::   ::::::
 386 heIleGlnPheThrGlnGluIleArgProThrValValLysGluAsnPro 402

1864 ...............CAAGTCAGCAGCGCGTATGCACAGGCAATAAGAGA 1898
                    ::::::::::::      ::::::      |||||
 403 AspLysGlyLeuIleGluIleThrLysIleIleGlyGluArgTrpArgGl 419

1899 AAGAAAGAGATGTGTTACTGCACACTGGTTAAACACAGTCTTAAAAAAAA 1948
     |         :::      |||::::::   :::|||      |||:::
 419 uLeuAspProAlaLysLysAlaGluTyrThrGluThrTyrLysLysArgL 436

1949 AAAAAATGGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTTCCCA 1998
       |||                        ::::::      ::::::|||
 436 euLys........................GluTrpGluSerCysTyrPro 444

1999 ............CCAGGAGGAAAGCCA 2013
                 |||:::|||:::|||
 445 AspGluAsnAspProAsnGlyAsnPro 453

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-258-442-13

seq_documentation_block:
; Sequence 13, Application US/08258442
; Patent No. 5670621
;  GENERAL INFORMATION:
;    APPLICANT:  Donahue, Brian A.
;    APPLICANT:  Toney, Jeffrey H.
;    APPLICANT:  Bruhn, Suzanne L.
;    APPLICANT:  Pil, Pieter M.
;    APPLICANT:  Brown, Steven
;    APPLICANT:  Kellett, Patti
;    APPLICANT:  Essigmann, John M.
;    APPLICANT:  Lippard, Stephen J.
;    TITLE OF INVENTION:  DNA Structure Specific Recognition
;    TITLE OF INVENTION:  Protein and Uses Therefor
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  2 Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,442
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/539,906
;      FILING DATE:  18-JUN-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4787AAA
```

```
; TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  617-861-6240
;    TELEFAX:  617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
;    LENGTH:  542 amino acids
;    TYPE:  amino acid
;    TOPOLOGY:  linear
; MOLECULE TYPE:  peptide
; ORIGINAL SOURCE:
;    ORGANISM:  Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
;    CLONE:  fractional yeast SSRP (fySSRP) (predicted)
; US-08-258-442-13

alignment_block:
US-09-664-641-10 x US-08-258-442-13

Align seq 1/1  to: US-08-258-442-13  from: 1  to 542

alignment_scores:
Percent Similarity:  45.972   Percent Identity:  20.236
Percent Quality:  184.00
Ratio:  0.786
                            Length:       509
                            Gaps:          23
```

```
832  TATAAGAGGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGTAGAAAATGA  881
       :|:|  | |||||||||||||||||||||||||||||  :|:||
4    PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGl  20

882  GGAACAAGATTCTCAGAAATGAGGGTAGTACAGATGACAAGTCAAGCCCTG  931
       |:||| |    :||||||| | |||||||||| :|||| :|:|||
20   nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln.  32

932  CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCAGCAG  969
       ||||||||| |||||||||||||||||||||||||
33   .........SerProGlnGlnGlnGlnInAsnValTyrPhe  43

970  .TTTTCACCTAAATCCAACACTGAAAAAATCTAAAGGGAAATAAAGTT  1016
       |||| ||||||| |||||  ||||||||||||||||| :|:
44   ProLeuProProGlnSerLeuThrGlnProThrSerGlnSerGlnGl  60

1017 TCATGATTCTTCACATTCACCGAAAAACAGGAGAGAAATTTAAACT  1066
       :|||| || :||| |||||||||| |||||| ||||:||
60   nGlnGlyHisTyrAlaAsnSerAsnSerAsnAsnSerAsnAsnValAsnV  77

1067 GGACCCCGGCCGGAAGTCCCACAGTTAGCTGCAGCAAAACCAGCGTGCT  1116
       :|:      |||||||||||| ||||||||||||||||||||||
77   alAsn........AlaLeuProGlnAspPheGlyTyrMetGlnThrThrGly  91

1117 CAGGGAAAGGAGCCTGGGTTGAATAAC  1143
       :|||||||                |||||||
92   SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnInPheSerGluPheTy  108

1144 ........TTGTGTTGCCAATGTCCACCCGTCCCAGTAACACATTTGCCCC  1186
                |||||||
108  ......SerAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.  122

1187 CTGAGGTCGGGGGT........................AAT  1203
       |||||||                                |||
122  roSerValThrGlyTyrThrGlyTyrAlaSerSerAsn...........  138

1204 TTAATGGGCTGCTGGACAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT  1253
       |||||||||:::
139  AsnValSerSerGlyAsnAsnSerThrSerSerAsnProThrGlnLeuAl  155

1254 AGCTACC...TGGAGTCCAGCTGTACCGACACTGAGGAATATTACTAATA  1300
       |||:|   :|||||||||:::
155  aAlaSerGlnLeuAsnProAlaThrAlaThrThrAlaAlaAlaAsnAsnA  172

1301 ATGCT........................GACAATCAGCAGAATGAAACCGG  1326
       |||                          ::||||
172  laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr  188
```

```
1327 ...CCAACAAATGTAGCACATATCTTACAGACTCTTTCAGGCACCTAGGAA  1373
       |||:|||:||:||           ::: |||:|:||||||   |||:::::
189  TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe  204

1374 A.....................................  1374
       :
204  rAlaGlyAlaSerAlaAlaGlyAlaAsnAlaAsnThrAlaThrHisProGlyLeul  221

1375 .......AATTTAGAACAGCAGGTGAATCAGCAGCAGGCAGGACAATACA  1416
       ||||:||:::         |||:||||  ||||||:||||
221  euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln...  234

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA  1466
       |||||||||||||||||||||||||||||||||||||||||
234  ........................................  234

1467 CATGTCTTACAGCAGCAGCAGGCCCACCAGCAGCAGCAGCAGCACCCGG  1516
       :::||:||||||||| :||| |||||||||||||||||||| ::::
235  .GlnMetGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGln  247

1517 TTTTACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAG  1566
       :|||||||||||||||||||||||||||:::||
247  ACAGCAGCAGCAGCAGCAGCAGCAGCA.  247

1517 ...CAGCAGATCTTCAGAGAACCTTACCCCAGCAGCAGCCGCGGCCCATTTC  1616
       :::||:||||:|||||||  |||| |||||||||||||||||| ::::
257  CAGGATCTTCAGAGAACCTTACCCCAGCAGCAGCCGCGGCCCA  257

1617 ACAGCAGCAGCAGCACAGCAGCAGCCA.  1644
       :|||||||||||||||||||||:::|||
274  nGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT  291

1645 .......................CCACCA  1650
       ..
291  hrGlnSerArgIleGluArgArgLysGlnLeuLeuLysGlyProLys  307

1651 TCGCCTCAG........................CAGCATCAGCT  1670
       :|||||||
308  ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe  324

1671 TTTTGGACATGATCCAGCACAGTGAGAATTCCAGAAGAAGAGCCTTCTTAATTGG  1720
       :||||||||||||||||||||||||||||:  :|| :::|
324  uLeuGlnGlnInPheProGluAlaLysValAlaProGluLeuSerGlyLeu  339

1721 GATGTGTTGTTTGCAATTGCGGATTATCCAGACAGCAGATGTCTGAT.  1764
       :|| |||  |||||||
340  ......AlaSerAlaAlaArgTrpLysGluLeuThrAspAspGlnLys  352

1765 .........AAGCAACTGCTGCCGCCACCTGAAAAGG.  1793
       ::||||||||:::
353  LysProPheTyrGluGluProGluThrPheLysTyrArgValVa  369

1794 AATCCAAGCACACATGGCGGCCGCCACTGTTGACCCC.  1824
       :|||||
369  lArgAspAlaTyrGlnLysThrLeuProProLysSerArgProSerGlyProp  386

1825 ....ACCTTCAGGAGTCGAATGCACGCACCACCTTCTTGTGAGT........  1863
       |||
386  helleGlnPheThrGlnGlnValIleAlaArgGluAsnPro  402

1864 .........CAAGTCAGCAGCGCGCTATGCACAGCAATAAGAGA  1898
       ::||
402  ......AspLysGlyLeuIleGluIleThrLysIleIleGluIleGlyLeuArgTrpArgGl  419

1899 AAGAAGAGATGTGTTACTCGACACTGGTTAAACACAGTCTTAAAAAAAA  1948
       |     :::||||||||:::|||
419  uLeuAspProLeuAlaLysLeuAlaGlyLeuTyrThrGlyThrTyrLysLysArgL  436

1949 AAAAAAATGTACCGGCCGCACCGAGCCCTTCACTTCCCAGTGGGCTTCCCA  1998
       |||                :::|:|||  |||
436  euLys..................GluTrpGluSerCysTyrPro  444
```

seq_documentation_block:

1999 ..............CCAGGAGGAAAGCCA 2013
445 AspGluAsnAspProAsnGlyAsnPro 453
Patent No. 5705334
seq_name: /cqn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-328-809-8

Sequence 8, Application US/08328809
GENERAL INFORMATION:
  APPLICANT: Lippard, Stephen J.
  APPLICANT: Essigmann, John M.
  APPLICANT: Donahue, Brian A.
  APPLICANT: Toney, Jeffrey H.
  APPLICANT: Bruhn, Suzanne L.
  APPLICANT: Pil, Pieter M.
  APPLICANT: Brown, Steven
  APPLICANT: Kellett, Patti
  TITLE OF INVENTION: Uses For DNA Structure-Specific
  TITLE OF INVENTION: Recognition Proteins
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibeault
    STREET: 53 State Street
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/328,809
  FILING DATE:
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Fenton, Gillian M.
    REGISTRATION NUMBER: 36,508
  REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617-248-7000
    TELEFAX: 617-248-7100
  INFORMATION FOR SEQ ID NO: 8:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 542 amino acids
      TYPE: amino acid
      TOPOLOGY: linear
    MOLECULE TYPE: peptide
    ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
    IMMEDIATE SOURCE:
      CLONE: fractional yeast SSRP (fySSRP) (predicted)
US-08-328-809-8

alignment_block:
US-09-664-641-10

Align seq 1/1 to: US-08-328-809-8 from: 1 to: 542

alignment_scores:
                Quality: 184.00   Length: 509
                  Ratio: 0.786   Gaps: 23
Percent Similarity: 45.972   Percent Identity: 20.236

alignment_block:
US-09-664-641-10 x US-08-328-809-8

882 GCAACAAGATCTCAGAATGACGGGTAGTACAGATGAGAAGTCAGCCCTG 931
    ::::||||::||::..::   ::||||
  4 PheGlnAlaGlnProGlnGlnGlnGlnGlnGlnGlnGl 20

832 TATGAAGAGGAAGAAGAGACCAAGAGAGGAGGAAGTAGAAAATGA 881
    :::::|| .:: ::::::::::::::::||||||||||||
  1   MetGlnGl 20

20 nGlnGlnGlnAlaProTyrGlnGlyHisPheGlnGln.............. 32
932 CCAGCTCTCAAGAAGGGTCTCCTTCAGGTGACCACCAG.......... 969
    ||||||:: ::|||||||| ::||||||
 33 ....SerProGlnGlnGlnGlnGlnAsnValTyrPhe 43

 44 ProLeuProProGlnSerLeuThrGlnProThrSerGlnLeuGlnGl 60
1017 TGATGATTCTTCAGATTCATCACGGAAAAAACAGGAGAGAAATTTAAACT 1066
     ::|||| ::|||| :::||| ::::||||||::||::::::: |||
 60 nGlnGlnGlnTyrAlaAsnSerAsnSerAsnSerAsnAsnAsnV 77

1067 GGACCCGCGCCGAAGTCCCACACAGTTAGCTGCGAAAACGCAGGGTGCCT 1116
     :::|||||||| ::|||||||||||| :::||
 77 alAsn......AlaLeuProGlnAspPheGlyTyrMetGlnGlnThrGly 91

1117 CAGGGAAAGGAGCCTGGGTTGATTAAC.............. 1143
     ::|||||||| ::||||
 92 SerGlyGlnAsnTyrProThrIleAsnGlnGlnGlnPheSerGluPheGly 108

1144 ...TTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTGCCCC 1186
    ::::||:::::::   :::|||| :::||||::::::||||
108 rAsnSerPheLeuSerHisLeuThrGlnLysGlnThrAsn.......P 122

1187 CTGAGGTCCGGGGT.................. 1203
    :::||||::::::
122 roSerValThrGlyThrGlyAlaSerSerAsnAsnAsnSerAsnAsnAsn 138

1204 TTAAATGGCTGCTGACAAAAACCTCCAAAGTTCTGAAAGATCAGAAATGAT 1253
     ::|| ::|||||| ::|||||||||||||||| :::||
138 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 155

1254 AGCTACC...TGGAGTCCAGCTGTGACGGGACTGAGGAAATTACTAATA 1300
     |||||| ::||||||||||:::::||||||||
155 aAlaSerGlnLeuAsnProAlaThrAlaThrAlaIaaAlaIaAsnAAsnA 172

1301 ATGCT.................GACATTCAGAGATGAAACGG... 1326
    ::|||| ::|||||||||||| :::||||
172 laAlaGlyProAlaSerTyrLeuSerGlnLeuProGlnValGlnArgTyr 188

1327 ...CCATCAAAATGTAGCACATATCTTACAGACTCTTTCAGCACTACCAA 1373
    ::::||::::|||| ::||||||||||:::   :::||||||
189 TyrProAsnAsnMet...AsnAlaLeuSerSerLeuLeuAspProSerSe 204

1374 A... 1374
204 rAlaGlyAsnAlaGlyAsnAlaAsnThrAlaThrHisProGlyLeuL 221

1375 ........AATTTAGAACAGCAGGTGAATCAGCCAGCAGGGACATACA 1416
    ::::::::::||||||:::   ::||||||||||||||||||||
221 euProProAsnLeuGlnProGlnLeuThrHisHisGlnGln........ 234

1417 AATGCCAATGCAGTGCTGTTTAGCCAAGTGAAAGTGACTCCAGAGACACA 1466
234 234

1467 CATGCTACAGCAGCACGGCCCAGCAGCGCAGCAGAAGTGGCAGCAGCGG 1516
    ::::|||||||||| ::::||||||||||||||||||||||||||
235 .GlnMetGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGln.......... 247

1517 TTTTACACCTTCAGCCCCACCAGATAATGCAGCTTCCAGCCAGCAGCAG 1566
    ::::||||||::   :::||||||||||||||||||||
248 ....LeuGlnLeuGlnLeuGln.......GlnLeuGlnGlnGln 257

1567 CAGGAGATCTCAGCACCTTACCCCCCAGCAGCGCGCCATCCATTTTC 1616
    ::::|||||||||||| :::|||||||||||||||| :::||||
258 HisGlnLeuGlnGlnGlnLeuGlnGlnGlnLeuHisHisHisLeuGl 274

1617 ACAGCAGCAGCAGCAGCAACAGCAGCCA.............. 1644
    :::||||||||||||||||||||||:::||
274 nGlnGlnGlnGlnGlnHisProValValLysLysLeuSerSerT 291

seq_name: /cgn2_6/ptcdata/2/iaa/PCTUS_COMB.pep:PCT-US92-11107-13

seq_documentation_block:
Sequence 13, Application PC/TUS9211107
; GENERAL INFORMATION:
~ ; APPLICANT: Donahue, Brian A.
~ ; APPLICANT: Toney, Jeffrey H.
~ ; APPLICANT: Bruhn, Suzanne L.
~ ; APPLICANT: Pil, Pieter M.
~ ; APPLICANT: Brown, Steven
~ ; APPLICANT: Kellett, Patti
~ ; APPLICANT: Essigmann, John M.
~ ; APPLICANT: Lippard, Stephen J.
~ ; TITLE OF INVENTION: DNA Structure Specific Recognition
~ ; TITLE OF INVENTION: Protein and Uses Therefor
~ ; NUMBER OF SEQUENCES: 13
~ ; CORRESPONDENCE ADDRESS:
~ ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
~ ; STREET: 2 Militia Drive
~ ; CITY: Lexington
~ ; STATE: MA
~ ; COUNTRY: USA
~ ; ZIP: 02173
~ ; COMPUTER READABLE FORM:
~ ; MEDIUM TYPE: Floppy disk
~ ; COMPUTER: IBM PC compatible
~ ; OPERATING SYSTEM: PC-DOS/MS-DOS
~ ; SOFTWARE: PatentIn Release #1.0, Version #1.25

~ CURRENT APPLICATION DATA:
~ ; APPLICATION NUMBER: PCT/US92/11107
~ ; FILING DATE: 19921218
~ ; CLASSIFICATION:
~ ; PRIOR APPLICATION DATA:
~ ; APPLICATION NUMBER: US 07/539,906
~ ; FILING DATE: 18-JUN-1990
~ ; ATTORNEY/AGENT INFORMATION:
~ ; NAME: Granahan, Patricia
~ ; REGISTRATION NUMBER: 32,227
~ ; REFERENCE/DOCKET NUMBER: MIT-4787AAA
~ ; TELECOMMUNICATION INFORMATION:
~ ; TELEPHONE: 617-861-6240
~ ; TELEFAX: 617-861-9540
~ ; INFORMATION FOR SEQ ID NO: 13:
~ ; SEQUENCE CHARACTERISTICS:
~ ; LENGTH: 542 amino acids
~ ; TYPE: AMINO ACID
~ ; TOPOLOGY: linear
~ ; MOLECULE TYPE: peptide
~ ; ORIGINAL SOURCE:
~ ; ORGANISM: Saccharomyces cerevisiae
~ ; IMMEDIATE SOURCE:
~ ; CLONE: fractional yeast SSRP (fySSRP) (predicted)
PCT-US92-11107-13

alignment_scores:
Quality: 184.00   Length: 509
Percent Similarity: Ratio: 0.786   Gaps: 23
45.972   Percent Identity: 20.236

alignment_block:
US-09-664-641-10 x PCT-US92-11107-13

Align seq 1/1 to: PCT-US92-11107-13   from: 1 to: 542

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-990-823-112

seq_documentation_block:
; Sequence 112, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; TITLE OF INVENTION: Immunostimulatory Peptides
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; CURRENT FILING DATE: 1997-12-15
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 112
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-08-990-823-112

alignment_scores:
Quality: 182.50
Ratio: 2.724
Percent Similarity: 41.875   Percent Identity: 31.875

alignment_block:
US-09-664-641-10/rev x US-08-990-823-112

Align seq 1/1   to: US-08-990-823-112   from: 1   to: 143

Length: 160
Gaps: 13

1688 GCTGGATCATGTCCAAAAAGCTGATGCTGCTCGAGGCGGATGGTGGCTG 1639
     :|:||||||||::||| |||||||||||||||||||||
   1 AlaGlyGlyCysGly.......AlaGlyGlyCysTyrThrThrGlyGlyCys 14

1638 CTGTTGCCTGCTGCTGCGC 1619
     |:::||||||||||||||:|:||||||||||:|||:||||||:T 1619
  14 SAlaCysCysCysAlaCysCysAlaCysGlyAlaGlyThrGlyAlaGlyThrC 31

1618 GTGAAAATGGATGCGGGGGGGGGTGCTGGGGGGTAAGGTTGCTGAGAGATCTGC 1569
     :||||||||||||:||| ||| ||||
  31 ysAlaGlyGlyCysAlaGly......GlyAlaGlyCys........Cys 42

1568 TGCTGC...TGCTGC. 1557
     |||||| ||||||
  43 CysCysGlyCysGlyThrThrGlyCysGlyGlyCysGlyGlyCys 59

1556 ....
     :||||||||||||:||| |||:||
  59 yGlyCysGlyCysGlyAlaAlaThrCysThrThrcGlyGlyCysGlyCys 76

1531 GCTGAAGGTGTAAAACCGGGTGCTGCTGCTGCTGCTGTGGGCCTGC.... 1485
     ||||| |||||:|||||||||
  76 ly.......ThrGlyGlyCysThrThrCysGlyGlyCysGlyAlaCysTyThr 89

403 AspLysGlyGlyLeuGluIleThrLysIleIleGlyLeuArgTrpArgGl 419
    :||::|| ||||
  419 AAGAAAGAGAGTGTTGTTACTGCACACTGGTTAAAACACAGTCTTAAAAAAA 1948
    ::| |||||||||:|||
  419 uLeuAspProAlaLysLysAlaLGluTyrThrGluThrTyrLysLysArgL 436

1949 AAAAAATGGTACCGCGCGCACCGACGCCTTCACTTCGCAGTGCCCTTCCCA 1998
     ||| ||:||| :||||:|||
  436 euLys..........GluTrpGluSerCysTyrPro 444

1999 .......CCAGGAGGAAACCA 2013
        :|||:|||:|||
  445 AspGluAsnAspProAsnGlyAsnPro 453

ACAGCAGCACAGCACAGCAAACAGCAGCCA.
:|||||||||||||||||||||||| .
nGlnGlnGlnGlnGlnGlnHisProValValLysLeuSerSerT 291

hrGlnSerArgIleGluArgAsnLysGlnLeuProLys 307
CAGCATCAGCT 1670
ArgProSerSerAlaTyrPheLeuPheSerMetSerIleArgAsnGluLe 324

seq_documentation_block:
Sequence 2, Application US/08916352
Patent No. 6161691
GENERAL INFORMATION:
APPLICANT: CHIRON CORPORATION
TITLE OF INVENTION: HUMAN POLYHOMEOTIC 1 (hph1) ACTS AS A
TITLE OF INVENTION: TUMOR SUPPRESSOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: 4560 HORTON STREET
CITY: EMERYVILLE
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,352
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: POTTER, JANE
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 1355.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-923-2707
TELEFAX: 510-655-3542
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1004 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

seq_name: /cqn2_6/ptodata/2/iaa/6B_COMB.pep:US-08-916-352-2

US-08-916-352-2

alignment_block:
US-09-664-641-10 x US-08-916-352-2

Percent Similarity: 44.589   Percent Identity: 21.934

alignment_scores:
Quality: 174.00   Length: 693
Ratio: 0.563   Gaps: 31

Align seq 1/1   to: US-08-916-352-2   from: 1   to: 1004

```
1597  ....CAGCCGCCCAICCAITTTCACCACGACGCACAACACGACC 1642
 447  aProGlnProPro..                          GlnValProProThrGlnGlnV 459

1643  CACCCACCAICCGCCTCAGCAGCAICCACGCGTTTTGGACAGAICGACCAGCAGIG 1692
 459  aProProSerGlnSerGlnGlnGlnAla.GlnProValValGlnPr 475

1693  GAGAITCCAGAAGAAGGCTTCTTAITGGGAIGTGIGTTTGCAATGCGG. 1741
 475  oMetLeuGlnSerSerPro.                    LeuSerLeuPro 486

1742                 AITAICAGAGCAGAGATTGCTGATA 1765
 486  roAspAlaAlaProLysProProIleProlleGlnSerLysProProVal 502

1766  AGCAACTGCTGGCCACCTGGAAAAGGAITAAICCAGGCACAATGCGGGCACT 1815
 503  AlaProIleLysProProGlnLeuGly.  AlaAlaLysMetSerAlaAl 518

1816  GTTGACCCCACCTTCACGAGTCGCAICGCACGCA.CCTTCTCGTGACAGTCA 1865
 518  aGlnGlnProProProHisIleProValGlnValGlyThrArgGlnP 535

1866  AGTCAGCAGCCGCGTAICGCACAGCAATAACGACAAAAGAGAGTGTGTTA 1915
 535  roGlyThrAlaGlnIleAlaGlnLeuGlyLeuAlaGlnLeuAlaLeuAlaAla 551

1916  CTGCACACT.....                      GGTTAAACACAGTCTTAAAAAAAA 1950
 552  ValProThrSerArgGlyMetProGlyThrValGlnSer.GlyGlnAlaH 568

1951  AAAATGTACCGCCGCACCGAGCCCTTCACTTCCCAGTGGCCTT. 1994
 568  iSerLeuAlaSerProProSerGlnAlaProGlyAlaLeuGlnGlu 584

1995  ..........CCCACCAGGACGCAAAACCAIGTTCACAGCAIAITATTTCGTGA 2038
 585  CysProProThrLeuAlaProGlyMetThrLeuAla........ 596

2039  CTGGAITTGTTGAIPAGTGACAGAGATGACCTAAAATTAATGGCTTAITTG 2088
 596  .....                                        596

2089  GCAGGTGCCAAAATAIACGGGTTAICTAIGCCCAGCAAACAGTCCTCAT 2138
 597  ....              ProValGlnGlyThrAlaHis 603

2139  CTGIAAAGAACCAAACTGCTTTAAAGTAIGCAAAAAGAGTGGAGGA 2188
 604  ValValLysGlyGly                ...                603

2189  TACCCTGTGTCAACGCCCAGTGGCTGGCTTCTCTGGGAAACTTT 2238
 614  oValValAlaGlnValProAlaIAlaPheTyrMetGlnSerValHisleuP 631

2239  GAGGC.                              ACTGAGGCAGATTCAGTAIAG 2264
 631  roGlyLysProGlnThrLeuAlaValLysArgLysSalAspSerGluGlu 647

2265  TCGCTACACGGCCATTCAGTCTGCCAGGATCCATTTGCCCCTACCACCATT 2314
 648  GluArgAspAspValSerThrLeuGlySerMetLeuProAlaLysAla.. 663

2315  TAGTTTAAAATCTTTAGAGITGGAGAGTTCCCTTAAAAGTGTCTGCA 2364
 664  .....                      SerProValAlaGluSerPro. 671

2365  GAGTGTTGTTGATGAGTATAAGACTACCTCCAAAACTAGAAGAATGAAGT 2414
 671  ysValMetAspGlu..LysSerSerLeuGlyLysAlaGlu.SerVa 686

2415  AGCTAAITGTCCAG..........CCTTCTTCCCAAAGAGCAGAATT.... 2451
                                                            2451
```

```
      ||||||||||::   ::           ::
 686  IAIaAsnValAsnAlaAasnThrProSerSerGluValAlaLeuThrp 703

2452  ....GAAGACGTTACCACCTCCCACT......................... 2472
 703  roAlaProSerValProProProThrLeuAlaMetValSerArgGlnMet 719

2473  ....                              AAAAGCTAACTCC 2486
 720  GlyAspSerLysProProGlnAlaIaIleValLysProGlnIleLeuThrHi 736

2487  AGAATTGACCCCTTTTGTGCTTTTCACTGGAITCCGAGCCGTTGTCCAGGTT. 2535
 736  sIleIleGluGlyPheVallleGlnGlyValAlaGluProPheProVal 753

2536  .....CAACAGTAITATTAAGAAG 2553
 753  IyCysSerGlnLeuLeuLysGlu 760
```

Quality: 173.00   Length: 97
Ratio: 3.681   Gaps: 8
Percent Similarity: 48.454   Percent Identity: 43.299

alignment_block:
US-09-664-641-10/rev x US-08-508-761B-22

Align seq 1/1   to: US-08-508-761B-22   from: 1   to: 120

1688 GCTGGATCATGTCCAAAAAGACTGATGCTGCTGAGGCGATGGT......GG 1645
     |||||| || |||||| ||| ::: |||
   21 AlaGlyCysCysThrGlyThrGlyThrGlyCysCysTh 37

1644 TGCCTCCTCTGTTCCTGCC........TGCTCCTCCTGTGAAAATGGAATGCGGCG 1601
     |||||| ||||||||| ||||||||||||||||||
   37 rGlyCysGlyCysCysGlyCysCys...ThrGlyCysGlyc 53

1600 GCTGCTGGGCGGCTAAGGTTCC...........TGAGAGATCGTGCTGC......... 1566
     |||||| ||||||| ||||||
   53 ysCysThrGlyThrGlyCysCysThrGlyAlaGlyCysCysThrGlyGly 69

1565 ...TGCTGC..........TGCTGCTGGAGCTGCATTATTCTGCTGGGG 1531
     |||||| |||||| |||||||
   70 GlyCysCysThrGlyGlyGlyCysCysThrGlyGlyGlyCysCysThrGl 86

1530 CTGCAAGGTGT...........AAAACCCGGCTTGCTGCTGC......TGCTGCT 1496
     ||||| ||||||
   86 yGlyGlyCysCysThrGlyThrGlyCysCysThrGlyGlyGlyCysCysT 103

1495 GCTGCGGCCTGCTGCTTCCTCCTGCTCTAGCATGTGTGTCTCTGGA 1455
     ||| |||||| ||| :::||||
  103 hrGlyAlaGlyCysCysThrGlyThrGlyCysCysThrGlyCysCysThrGly 116

seq_documentation_block:
Sequence 20, Application US/09625188
Patent No. 6307037
GENERAL INFORMATION:
APPLICANT: NO. 6307037artis AG
TITLE OF INVENTION: Fungal Target Genes and Methods
FILE REFERENCE: PH/5-31285pl
CURRENT APPLICATION NUMBER: US/09/625,188
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 20
LENGTH: 729
TYPE: PRT
ORGANISM: Ashbya gossypii

seq_name: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:US-09-625-188-20
US-09-625-188-20

alignment_block:
US-09-664-641-10 x US-09-625-188-20

Align seq 1/1   to: US-09-625-188-20   from: 1   to: 729

Percent Similarity: 50.150   Percent Identity: 25.225

alignment_scores:
Quality: 173.00   Length: 333
Ratio: 1.036   Gaps: 15
Percent Similarity: 48.454   Percent Identity: 43.299

841 GAAGAAGAAGAAGAGAAGGAGGAGGAAGTAGAAAATGAGGAACAAGA 890
    ||| ::||| ::::::||:::::|||
227 GluValValAspGluLeuThrGlnGlnIleProPheGlnThrGlnMe 243

891 TTCTCAGAATGAGGGGTAGTACAGATGAGAAGTCAAGCCCTGCCAGCTCTC 940
    |||||| |||||:::::|| :::||
243 tGlnMetHisGlyGlySerProAlaProThrHisValThrAlaTyrG 260

941 AAGAAGGGTCTCCTTCAGGTGACCAGCAGTTTCACCTAAAATCCAACACT 990
    |||:: |||||||

OM Of: US-09-664-641-10 to: A_Geneseq_032802.*   out_format : pfs

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Date: Jun 4, 2002 5:32 PM

Command line parameters:
-MODEL-frame+p2p.model
-Q=/cgn2.1/USPTO_spool/US09664641/runat_04062002_110932_9158/app_query.fasta_1.3697
-DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT=0.000
-QGAPOP-4.500 -QGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500
-DHLOP-6.000 -DHLEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN-200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTPMT=pfs
-NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000000
-USER-US09664641 @CGN1_1_206 -NCPU-6 -ICPU-3 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT-30 -NO_XLPXV -WAIT -THREADS-1

Search information block:
Query: US-09-664-641-10
Query length: 3580
Database: A_Geneseq_032802.*
Database sequences: 747574
Database length: 111073796
Search time (sec): 168.130000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	! Documentation
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU278822	+	4791.00	8230.72	0.0	91
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB71695	+	852.00	1419.51	3.5e-72	17
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27997	+	319.00	535.15	1.1e-21	10
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB71695	+	294.50	488.80	3.1e-19	14
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AA005204	+	263.50	405.29	1.3e-15	15
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY03183	+	249.00	405.29	4.3e-15	61
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23599	+	248.00	395.52	1.7e-14	39
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG19008	-	242.00	404.43	2.5e-14	87
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23600	-	238.00	360.00	3.9e-13	16
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAU07343	-	225.50	348.20	3.8e-12	77
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69280	+	225.50	334.19	1.5e-11	11
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63516	+	220.50	351.12	3.3e-12	89
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAY27446	-	220.00	366.15	1.2e-11	16
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY60558	+	216.00	336.94	2.4e-11	51
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61245	+	216.00	358.24	9.1e-12	90
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAY27784	-	215.50	323.22	6.3e-11	11
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY96255	+	214.00	323.22	6.3e-11	11
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAY58500	+	214.00	323.22	6.3e-11	11
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB62331	+	213.50	313.79	3.4e-10	24
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB63057	+	213.50	304.32	3.4e-10	24
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB68996	+	208.50	338.84	8.1e-11	12
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27074	-	206.50	338.03	8.4e-11	13
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23598	-	206.50	338.03	8.4e-11	13
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67633	-	206.50	313.16	2.9e-10	92
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG29284	-	205.00	310.45	4.0e-10	93
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAY54320	+	205.00	292.02	2.0e-09	20
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB67519	+	200.00	292.02	2.0e-09	76
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM41453	+	199.00	311.22	8.8e-10	38
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY54319	+	199.00	289.97	2.5e-09	20
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB58382	+	198.50	284.63	3.9e-09	35
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG17148	+	198.00	284.63	3.9e-09	27
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB72020	+	197.50	302.47	1.7e-09	63
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70437	-	197.00	280.73	6.4e-09	32
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27789	-	196.50	324.61	5.4e-10	96
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB84634	+	195.50	278.47	7.1e-09	31
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB70437	+	195.50	325.34	8.1e-10	69
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27787	+	194.50	312.77	1.5e-09	69
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB62018	+	194.50	292.78	4.0e-09	18
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27787	+	193.50	288.23	5.8e-09	90
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR38746	+	193.00	296.62	4.1e-09	54

seq_documentation_block:

ID   AAU27822 standard; Protein; 911 AA.
XX
AC   AAU27822;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human full-length polypeptide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO20016834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0618807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   N-PSDB; AAS44722.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection.
XX
PS   Claim 10; SEQ ID NO 319; 153pp; English.
XX
CC   Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC   contig polypeptides of the invention. The proteins and their associated
CC   DNA sequences are useful for the treatment, diagnosis and prevention of
CC   various types of disorder in a mammalian subject such as a human, dog,
CC   monkey, mouse, hamster or rat. The disorders include cancers such as
CC   leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC   multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to manipulate stem cells

seq_name: /SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27822

/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAM39667	+	193.00	295.85	4.3e-09
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27786	-	192.00	318.08	1.6e-09
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB30808	+	192.00	303.00	3.4e-09
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB61004	+	192.00	289.00	6.8e-09
/SIDS5/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB64877	+	192.00	272.87	1.5e-08

Same (handwritten annotation)

CC　in culture to give rise to neuroepithelial cells that can be used to
CC　augment or replace cells damaged by illness, accidental damage or genetic
CC　disorders. The sequences may also be used for regeneration of bone,
CC　cartilage, tendons and ligaments and in tissue repair and burn healing.
CC　Note: Some sequences for this patent did not form part of the printed
CC　specification, but were obtained in electronic format directly from WIPO
CC　at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ　Sequence　911 AA;

alignment_scores:
　　Quality: 4791.00　　　　　　　Length:　911
　　Ratio:　5.259　　　　　　　　 Gaps:　0
　　Percent Similarity: 100.000　 Percent Identity: 100.000

alignment_block:
us-09-664-641-10 x AAU27822

Align seq 1/1　to: AAU27822　　from: 1　to: 911

```
 373 ATGGTTTTCTCAGAATCATGTCAGATTTTTTTTGGAAACACTCCTGC 422
     |||||||||||||||||||||||||||||||||||||||||||||||
   1 MetValPheLeuGlnHisValArgPhePheLeuGlnSerLeuProAl 17

 423 CTTTCTCAGGGTGGTTGGATCAAGCTGGACCTCTTCGTTGAGAGTCTTCCAG 472
     |||||||||||||||||||||||||||||||||||||||||||||
  17 aPheLeuArgValLeuLeuGlnAlaGlyAlaLeuCysTrpSerLeuProG 34

 473 AGCTCTCCAAGGAGGAGGTAGGGAAGGAGACCTTGTCCAGCAGAAGTTGGG 522
     ||||||||||||||||||||||||||||||||||||||||||||||||
  34 luLeuSerGlnGluValGlyLysGluThrLeuSerSerSerArgTrpG 50

 523 AACGACAGAGATCATCTGCCCTTCTTCTGACCCGGTATTGATGCAGGCTGA 572
     ||||||||||||||||||||||||||||||||||||||||||||||||
  50 luGluSerGlyAspHisLeuProSerSerAspProValLeuMetGlnAlaGl 67

 573 GGCCTCTGTTGTTAATGTCTGGTGTCATCTGAAGACAGAAGTCGCCTGT 622
     |||||||||||||||||||||||||||||||||||||||||||||||
  67 uAlaSerValValMetCysTrpValSerSerGluAspArgSerAlaLeuT 84

 623 GGGCTTTGGTTTACGGTTCTTATGGGGGAGATTGCCAGCTAACCTCAATAAG 672
     |||||||||||||||||||||||||||||||||||||||||||||||
  84 rpAlaLeuValThrPheTyrGlyGlyAspCysGlnLeuThrLeuAsnLys 100

 673 AAATGACACGCATTTGATTTGTTCCAGACCCAAAGCGCGGAAATACGAATG 722
     |||||||||||||||||||||||||||||||||||||||||||||||
 100 LysCysThrHisLeuIleValProGluProGlyGlyLysTyrGluCy 117

 723 TGCTTTAAAGGGAGCAAGTATTATTAAAATTGTGACTCCTGACTGGGTTCTGG 772
     ||||||||||||||||||||||||||||||||||||||||||||||
 117 sAlaLeuLysArgAlaSerIleLysIleValThrProAspTrpValLeuA 134

 773 ATTGGCTGTATCAGAGAAAACCAAAAAGGACGACGAAGCCATTTTATCATCCTCGT 822
     |||||||||||||||||||||||||||||||||||||||||||||||
 134 spCysValSerGluLysThrLysLysAspGlnAlaPheTyrHisProArg 150

 823 CTGATTANTTATGAAGAGGAAGACGAAGCAAGAGCAGAGAGGAGAGT 872
     ||||||||||||||||||||||||||||||||||||||||||||||
 150 LeuIleLeuTyrGluGluGluGluGluGluGluGluGluGluVa 167

 873 AGAAAAATGAGGAGCAAAGATTCTCAGAATAGGGTAGTACAGATGAGAGT 922
     |||||||||||||||||||||||||||||||||||||||||||||||
 167 lGluAsnGluGluGlnAspSerGlnAsnGlySerThrAspGluLysS 184

 923 CAAGCCCTGCCAGCTCTCAAGAGAGGTCTCTCCTTCAGGTGACCAGCAGTTT 972
     |||||||||||||||||||||||||||||||||||||||||||||
 184 erSerProAlaSerSerGlnGluGlySerProSerGlnAspGlnGlnPhe 200

 973 TCACCTAAAAATCCAACACTGAAAAAATCTAAAGGGGGAATTAATTGTTTGATGA 1022
     |||||||||||||||||||||||||||||||||||||||||||||||
 201 SerProLysSerAsnThrGluLysSerLysGlyGluLeuMetPheAspAs 217
```

```
1023 TTCTTCAGATTCATCACCGGAAAAACGAGGAGAGAGAAAATTTAAACTGGACC 1072
     ||||||||||||||||||||||||||||||||||||||||||||||
 217 pSerSerAspSerSerProGluLysGlnGluArgAsnLeuAsnTrpThrP 234

1073 CGGCCGAAGTCCCACAGTTAGCTGCAGCAAAAACCAGGCTCCCTCAGGGA 1122
     ||||||||||||||||||||||||||||||||||||||||||||||
 234 roAlaGluValProGlnLeuAlaAlaAlaLysThrArgLeuProGlnGly 250

1123 AAGGAGGCCTGGGTTGGATTAACTTGTGTGCCAATGTCCCACCCGTCCCAGG 1172
     |||||||||||||||||||||||||||||||||||||||||||||||
 251 LysGluProGlyLeuIleLeuAsnLeuCysValAlaAsnValProValProGl 267

1173 TAACATTTTGCCCCCTGAGGTCCGGGGTAATTTTAATTGGCTGTGACAAA 1222
     ||||||||||||||||||||||||||||||||||||||||||||||
 267 yAsnIleLeuProProGluValArgGlyAsnLeuMetAlaAlaLeuGlnMetAs 284

1223 ACCTCCCAAAGTTCTGAAACGATTACAGAAAATGATTAGCTTACCTGGAGTCCAGCT 1272
     ||||||||||||||||||||||||||||||||||||||||||||||
 284 snLeuGlnSerSerGluArgSerGluMetIleAlaLeuThrTrpSerProAla 300

1273 GTACGGACACTCAGGGAATATTACTAATAATAAAAAATCCTGACATTCACGACATGAA 1322
     |||||||||||||||||||||||||||||||||||||||||||||||
 301 ValArgThrLeuArgGluAsnIleThrAsnAsnAlaAspSerIleGlnMetAs 317

1323 CCCGCCCATCAAAATGTAGACACATATCTTACAGACTCTTTCACCACCTACGA 1372
     ||||||||||||||||||||||||||||||||||||||||||||||
 317 nArgProSerAsnAlaAlaHisIleLeuGlnThrLeuSerProThrThr 334

1373 AAAATTTAGAACAGCAGGGTGAATCACAGCCCAGGCAGGACATACAAAATGCC 1422
     |||||||||||||||||||||||||||||||||||||||||||||||
 334 ysAsnLeuGluGlnGlnValAsnHisSerGlnGlnHisThrAsnAla 350

1423 AAATGCAGTGCTGTTTAGGCCAAGTGAAAGTGACTCCAGAGACACACATGCT 1472
     ||||||||||||||||||||||||||||||||||||||||||||||
 351 AsnAlaValLeuPheSerGlnValLysValThrProGluThrHisMetLeu 367

1473 ACAGCACAGCAGAGCAGCGCCCAGCAGCAGCAGCAGCAGCACCCGGTTTAC 1522
     |||||||||||||||||||||||||||||||||||||||||||||||
 367 uGlnGlnGlnGlnAlaGlnGlnGlnGlnHisProValLeuH 384

1523 ACCTTCAGCCCCCAGCAGATAATCCAGCTCCAGCACCAGCAGCAGCAG 1572
     ||||||||||||||||||||||||||||||||||||||||||||||
 384 isLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGln 400

1573 ATCTTCCAGCAACCTTACCCCCAGCACCGCGCCATCCATTTTCACAGCA 1622
     ||||||||||||||||||||||||||||||||||||||||||||||
 401 IleSerGlnGlnProTyrProGlnHisProHisSerSerGlnI 417

1623 GCAGCACAGCAAACAGCAGCCACCACCATCGCCTCCAGCAGCATCAGCTTT 1672
     |||||||||||||||||||||||||||||||||||||||||||||||
 417 nGlnGlnGlnGlnGlnProProProSerProGlnHisGlnLeuP 434

1673 TTGGACATGATCCAGCCAGTGGCAGATTCCAGAGCAAGAGGCTTCTTTATTGGGA 1722
     ||||||||||||||||||||||||||||||||||||||||||||||
 434 heGlyHisAspProAlaValGlnLeuProGluGlnGlyPheLeuLeuGly 450

1723 TGTGTGTTTGCAATTGCGGATATTCCAGAGACGAGATGTCTGATAAGCAACT 1772
     |||||||||||||||||||||||||||||||||||||||||||||||
 451 CysValPheAlaIleAlaAspIleProGluThrMetSerAspLysGlnLeu 467

1773 GCTGGCCACCTGGAAAAAGGATAATCCAGGCACATGGCGGGCACTGTTGACC 1822
     ||||||||||||||||||||||||||||||||||||||||||||||
 467 uLeuAlaThrTrpLysArgIleIleGlnAlaHisGlyGlyThrValAspP 484

1823 CCACCTTCACCAGTCGGATCCACGCCACCTTCTCTGAGAGTCAAGTCAA 1872
     |||||||||||||||||||||||||||||||||||||||||||||||
 484 roThrPheThrSerArgCysThrHisLeuLeuCysGluSerGlnValSer 500

1873 AGCCCGTATGCACGGCAATAGCAGAAAGAAAGAGACGTGTTACTGCACA 1922
     |||||||||||||||||||||||||||||||||||||||||||||||
 501 SerAlaArgAlaGlnHisAlaIleArgGluArgGlyCysValThrAlaHi 517

1923 CTGGTTAAACACAGTCTTAAAAAAAAAAAAAATGTACCGCCGCACCGAG 1972
```

```
 517 ||||||||||||||||||||||||||||||||||||||||||||||||||||
     sIrpLeuAsnThrValLeuLysLysLysMetValProProHisArgA 534

1973 CCCTTCACTTCCCAGTCGCCTTCCACCAGGAGAAAGCCATGTTCACAG
     |||||||||||||||||||||||||||||||||||||||||||||
 534 laLeuHisPheProValAlaPheProGlyGlyLysProCysSerGln 550

2023 CATATTATTTCTGTGACTGGATTTGTTGATAGTGACAGAGATGACCTAAA
     ||||||||||||||||||||||||||||||||||||||||||||||
 551 HisIleIleSerValThrGlyPheValAspSerAspArgAspAspLeuLys 567

2073 ATTAATGGCTTATTTGGCAGGTGCCAAATATACGGGTTATCTATGCGGCA
     |||||||||||||||||||||||||||||||||||||||||||||||
 567 sLeuMetAlaTyrLeuAlaGlyAlaLysTyrThrGlyTyrLeuCysGlyA 584

2123 GCAACACACTCCTCANTCTGTAAACAACCAACTCGTTTTAAAGTATGAAAAA
     |||||||||||||||||||||||||||||||||||||||||||||
 584 erAsnThrValThrLeuCysLysGluProThrGlyLeuLysTyrGluLys 600

2173 GCCAAAGAGTGGAGGATACCCCTGTGTCAAAGCCCACTGGGCTTGGCGACAT
     |||||||||||||||||||||||||||||||||||||||||||||
 601 AlaLysGluTrpArgIleProCysValAsnAlaGluThrGlyLeuGlyAspIl 617

2223 TCTTCTGGGAAACTTTGAGGCACTGAGGCAGAGTTCAGTAGTATAGTCGCTACA
     |||||||||||||||||||||||||||||||||||||||||||||
 617 eLeuLeuGlyAsnPheGluAlaLeuArgGlnIleGlnInTyrSerArgTyrY 634

2273 CGGCATTCAGTCTCCCAGGATCCACCATTGCCCCTACCCACCATTTACTTTTA
     |||||||||||||||||||||||||||||||||||||||||||||
 634 hrAlaPheSerLeuGlnAspProPheAlaProThrGlnHisLeuValLeu 650

2323 AATCTTTTAGATGCTGTGGAGAGTTCCTTAAAAGTGTCTGCAGAGTTGTT
     |||||||||||||||||||||||||||||||||||||||||||||
 651 AsnLeuLeuAspAlaValTrpArgValProLeuLysValSerAlaGluLeu 667

2373 GATGAGTATAAGACTACCTCCCAAACTGAAACAGAATGAAGTAGCTAATG
     |||||||||||||||||||||||||||||||||||||||||||||
 667 uMetSerIleLeuArgLeuProProLysGlyLysGlnAsnGluValAlaAsnV 684

2423 TCCAGCCTTCTCCCAAAGACCCAGAATTCA-.-ACGTACCACCTCCCACT
     ||||||||||||||||||||||||||||||||||||||
 684 alGlnProSerSerGlnArgAlaAn:GllSGluAss:ValProProThr 700

2473 AAAAGCTAACTCCAGANTTGACCCCCTTTTGTGCTTTTCACTGGATTCGA
     |||||||||||||||||||||||||||||||||||||||||||||
 701 LysLysLeuThrProGluLeuThrProValLeuPheThrGlyPheGl 717

2523 GCCTGTCCCAGGTTCACACAGTATATTAAGNAGTTCTACATTCTTGGTGGAG
     |||||||||||||||||||||||||||||||||||||||||||||
 717 uProValGlnValGlnGlnTyrIleLysLysLeuTyrIleLeuGlyG 734

2573 AGGTTCGCGAGTGCTCCACAGAAGTGCACACACCTCATTGGCCACCAAAGTG
     |||||||||||||||||||||||||||||||||||||||||||||
 734 luValAlaGluSerAlaGlnLysCysThrGluGlyHisIleLeuVa 750

2623 ACTCGCACCGTGAAGTCCTGACGGCGCCGCANTTTCTGTCGTGAAGCACATAGT
     |||||||||||||||||||||||||||||||||||||||||||||
 750 ThrArgThrValLysPheLeuThrAlaIleSerValValLysHisIleVa 767

2673 GACGCGCCAGTGGCTCGGAGAATGCCTTCAGGTGTCAGAAGTTCATTGATG
     |||||||||||||||||||||||||||||||||||||||||||||
 767 lThrProGluTrpLeuGluGlyCysPheArgCysSerGlnLysPheIleAspG 784

2723 AGCAGAACTACTACTTCCCGACAGTCCTCAGGCGCAGAGTACTATTTCTCTTTTC
     |||||||||||||||||||||||||||||||||||||||||||||
 784 luGlnAsnTyrTyrLeuArgGlnSerSerArgAlaGluLeuValLeuPheSerPhe 800

2773 AGCTTGGAAGANNCCTTAAAANGGGGCACAGTTTCTCCCACTCTTTAAGGC
     |||||||||||||||||||||||||||||||||||||||||||||
 801 SerLeuGluGluSerLeuLysArgAlaHisValSerProLeuPheLysAl 817

2823 AAAATATTTTACNCACACCTGGAATCGNCCAAGTCTTTCCACTATGA
     |||||||||||||||||||||||||||||||||||||
     817
```

```
alignment_scores:
       Quality:    852.00        Length:     848
         Ratio:     1.906          Gaps:      24
Percent Similarity:  52.712   Percent Identity: 28.774

alignment_block:
us-09-664-641-10 x AHB71695

Align seg 1/1   to: AHB71695   from: 1  to: 1798
```

```
 925 AGCCCTGCCAGCTCTCAAGAAGGGGTCTCCTTCAGGTGACCAGCAGTTTTC 974
     ||||||        :::::||||          ||||||||||||||
 974 SerProIleGlnGlnPhrProGlnGlnGlnGlnGlyAlaGlyProGlnGlnGlnGlnGlnSe 990

 975 ACCTAAATCCAACACTGAAAAATCTAAAGGGCAATTAATCTTTGATCATT 1024
     :::||||||||||||                 |||||           ||||
 990 rProGlnGlnGlnProPheAsnProAsn...................AspProA 1002

1025 CTTCACATTCATCACCGAAAAAACAGAGCAGAAATTTAAACTGGACCCCG 1074
     |||              ||||||      |::::|||||
1002 snGlnGlnMetLeuMetGlnArgGlnGlnLeuArgGlnValGlnGlnIlePro 1018

1075 GCCGAA.............GTCCCACAGTTAACTGCAGCAAAACGAGGCTGCC 1114
     |||                      ::::|||||||        ||||
1018 eThrGlyProAsnGlyValGlnLeuIleGluGlnGlnThrIleValGlnGlnP 1052

1116 TCAGGCAAACGACGAGCCTTGGCTTGCATT................. 1140
     |||        |||||
1052 roGlnGlnGlnSerAlaProGlyThrProThrProGlyValAlaGlyGly 1068

1177 ATTTTGCCCCTGAGGTCCGGGTAATTAATGGCTGCTGGGACAAAACCT 1226
     |||||||::::::|||                   ||||||
1068 .......GlyAsnAlaAsnIleMetThrGlnThrIle 1081

1227 C....CAAACTTCTGAAAGATTCAGAAATCTAAATAGCTACCTGGAGTCCAGCTG 1273
     :::||||||                ||||||
1081 uValMetThrSerThrThrProAspGlyValGlnProGlnSerProGlnG 1098

1274 TACGGCACACTCAGGAAATATTACTAATAATGCTGACATTCAGCAGATGAAC 1323
     ||||||||                   :::|||||
1098 lnMetAsnLeuLysThrLysThrAlaLeuAlaAsnMetLeuSerAsnArg 1114

1324 CGGGCATCAAATGTAGCACAATANCTTACNGAACTCTTTCAGCACACTTACGAA 1373
     |||::|||                   |||
1114 LeuGlyAsnAsnGlyAlaGlnThrProGln.............. 1124

1374 AAAATTTAGACAGCAGCAGGTCGAATCACACCCACGAGGACAATACAAAATGCCA 1423
     ||||||                  ||||||
1124 .......GlnGlnIleIleGlnLeuProAspGlyValGlnProGlnG 1138

1424 ATGGCAGTTGCTCGTTTAGCCAAGTGAAAAGTGACT..........CCA 1458
     :::||||||                           |||
1138 lnGlnIleGlnIleValGlnGlnValValThrGlyAlaIleGlnGlnGln 1154

1459 GAGACACACATGCTACAGCAGCAGGCCAGCGAGCGAGCAGCAGCA 1508
     |||||||||||||||||||                ||||
1154 GlnGlnIleLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl 1171

1509 G.............CACCCGGTTTTACACCTTCAGCCC........... 1533
     |||                    |||
1171 nMetValValGlnGlnGlnLeuProProGlnGlnHisValGluProPros 1188

1188 erAlaAlaGlyAlaLeuAargMetMetGlyGlnGlnHisAsnAlaThrAla 1204

1533 ..............................  1533

1534 ..............................  1556

1188 ..............CAGCAGAATAATGCAGTCCAGCA 1204
                    :::|||
                    CAGCAGAATAATGCAGTCCAGCA 1556
```

```
1205 GlyValProGlyProProArgThrGlnGlnGlnGlnLeuLeuMetLeuGlnGl 1221
     |||::::::::|||||||::::::::|||||||
1557 GCAGCAGCAGCAGCAGATCTCCAGCAA................... 1584

1221 nGlnGlnGlnGlnInLeuLeuGlnGlnGlnGlnGlnMetArgArgValValA 1238
     ||||||                     ||||||
1585 .......CCTTAGCCCCAGCAGCCGCGGCGCCATCATT.......... 1614

1238 laLeuProGlnGlnGlnInAlaProGlnGlnHisGlyTyrIleg 1254
     ||||||              |||||
1615 ...............TCACAGCAGCAGCAGCAA.......... 1635

1255 IleValAlaAlaProThrGlnGlnSerProGlnGlnGlnMetAlaVa 1271
     |||||||||||              |||
1636 .......CAGCAGCCACCACATCCGCCCTCAGCAGCCATCAGCTT 1671

1271 lGlyValGlyValGlyValGlnAsnArgThrProGlnHisGlyTyrIleg 1288
     ||||||                   :::|||
1672 TTTGGACATGATCCAGCAGTGGAGATTCCAGAGAGAAGGCTTCTTNNTTGGG 1721

1288 lnThrArgProGlyGlnGlnInProIlePromet Pro..... 1301
     ||||||               |||||
1722 ATGTGTGTTTGCAATTGCGGATTATCCAGAGAGATGTCTGATAAGCAAC 1771

1302 TyrGlyHisAsnProAsnLeuLysLeuProAlaAspLeuPheLeuValGl 1318
     ::::::|||||::::|||          ||||||
1772 TGCTGGCCACCTGGANAAGGATAANTCCAGGCACACATGGCGCCACTGTTGAC 1821

1318 yCysThrPheTyrIleValGluTyrAspGluThrAspGlyAspGlu.... 1333
     :::|||
1822 CCCAGCTTCACGAGTCGAATGCAGCCACCTTCTGTGAGAGTCAAGTCAG 1871

1334 .LeuProIleTrpLeuAspThrIleArgGlnPheGlyGlyAspIleGlu 1349
     ||||                      :::|||
1872 CAGCGGCGTATGCACAGCCAATAAGAGAAGAAAGAGAAGTGTTTACTGCAC 1921

1350 ArgValTyr(,sProArgValThrHisValIleCysSerArgThrGlnAspHi 1366
     ::::|||             ||||||
1922 ACTGGTTAAACACAGTCTTAAAAAAAAAAAAATGTACGGCGCGCACCGA 1971

1366 sGlyValValMetGlnHisAlaLeuArgAspLysArgCysValThrAlaT 1383
     ||||||                  :::|||
1972 GCCCTTCACTTCCCCAGTGGCCTTCCCACCAGA................. 2012

1383 yrTrpLeuSerAspIleCysSerLeuTyrSerArgGlnLeuMetProProTrpGln 1399
     |||||                        ||||||
2013 ATGTTCACGACATATTATTTCTGTGACTGCAGATTCGTTGATATGACAGAG 2062

1400 ProLeuHisSerLeuPro......PhePtoSerGlnPheGlyTyrArgLysPr 1414
     ||||||                          ||||||
2063 ATGACCTAAAATTAATTAAGGCTTAATTGGCAGGTGCCAATATACGGGTTAT 2112

1414 oLeuGluArgTyrIleIlecThrSerGluGlyPheGluGlyGluGluValV 1431
     :::::|||          ||||||
2113 CTATGCCGCAGCACACAGTCCTCATCTGTAAAGAACCAACTGGTTTAAA 2162

1431 alArgLeuGlnGlnMetAlaGluGluCysGlyAlaLeuIleTyrPheSerTyr 1447
     ||||||                       :::|||
2163 GTATGAAAAAAGGCCAAAGAGTGGAAGGATATACCCTGTGTCAACGCCCCAGTGGC 2212

1448 LeuSer(/sValAsnThrValValValCysGlnLeuGlyLysAsnGly 1464
     ||                    ||||||
2213 TTGGCGACAATTCTCTGGGAAACTTTGAGGCACTGAGGCAGAGTTCAGTTAT 2262

1464 sPheAsnAlaAlaLysGluTrpAsnIleProMetValAsnAlaLeuTrpL 1481
     ::::::||              |||
2263 AGTCGCTACACGGCATTCAGTCGTGCAGGAGATCAATTTGCCCTACCCAGCA 2312

1481 euSerAspValCysIleGlyAsnLeuSerGlyTyrGluSerGlnTyrGluAsn 1497
     ::::::|||
                                                2312
```

```
1498 ProLysTyrGlnGlnTyrAsnLeuValAlaProPheArgIleGluCysAs 1514
     |||||||
2313 TTTAGTTTTAAATCTTTTAAATGATTGGGAAGTTCCCTTAAAAGTGTCTG 2362
1514 nLeuValAlaHisLeuLeuThrAlaTrpLysAlaProIleAsnLeuThrG 1531
     ::|||||| |||:||:|| |::|||||| :::||||||:
2363 CACAGTTCGTTCGATCGAGTATTAAACACTA.................. 2388
1531 lnGluAlaHisGluArgValLysArgHisLeuSerAspProTyrGlyAsn 1547
     :::||:
2389 CCTCCCAAACTGAAACACAGAATAGAGAGTAGCTAATGTCCCAGCCCTTCTTCCCA 2438
1548 GluGlnLysSerLysArgGlnLysMetAsnAlaPheGlnGlnGluGlnIle 1564
     |||:: :||||||:|||:::|||||||:
2439 AAGAGCCAGAATTGAAGACGTACCAACCTCCCACTAAAAGCTAACTCCAG 2488
1564 uProGluGlnIleValCysValGluTyrProThrThrThrLysProProCol 1581
     |||:
2489 AATTGACCCCTTTTGCTGCTTTCACTGAAGCCTCTGCCCAGGTTCAA 2538
1581 ysValIle.......................PheSerGlnValAlaAspAla 1590
     |||:|||
2539 CAGTATATTAAGAAG...CTCTACATTCTTGGTTGGGAGGGTTGGGGAGTC 2585
1591 GluAlaLeuLysAlaValLeuIleLeuGlyCylyIleValAlapSe 1607
     :::||| ||||:||
2586 TGCACAGAAGTGCACACCTCATTGCCGCCAAAGTGACTCGCACCGTGA 2635
1607 rProAlaAspAlaThrHisLeuValMetThrArgGluSerArgThrCysL 1624
     ::|||||||||||||||
2636 AGTTCCTGACGCGCCATTTCTCGTGAAGCACACATAGTGACCCAGAGTGG 2685
1624 ysLeuIleLnAlacyscysHisValAspTyrValLeuLysSerSerTrp 1640
     :::|||
2686 CTGGAAGAATGCTTCAGGTGGTGTCAGGAGTTCAATGATGAGGAGAACTACAT 2735
1641 IleAlaAspSerAlaLysAlaGlyValGlySPheValProThrAspProTyrAr 1657
     :::||||
2736 TCTCCCAGATGCTGAGCGAGAGTACTTTTCTCTTTCACGTTGGAAGAAT 2785
1657 GluIleGlnHisIleProValAspGluAsnLeuPheAsnLeuAsnThrV 1674
     :::||| |||
2786 CCTTAAAACGGCCACACGTTTTCTCCACTCTTTAGCGAAAATATTTTAC 2835
1674 alLeuCysAlaProThrArgSerThrLeuPheAlaGlyLysTyrPheHis 1690
     :::||||||||||:::
2836 ATCACACCTGGAATCTGCCCAAGTCTTTCCACTAGAAGGCAATCGTAGA 2885
1691 ValThrProAspValPhePrpAlaArgGluGluIleIleArgMetIleGl 1707
     :::||:||
2886 GTGTGTCAGGAGGAGAAGTGTTATCCAGGACCACCTTCCGGAAGCTCA 2935
1707 uSerSerGlyLysValGluProLysArgArgSerGlyAlaSerValA 1724
     :::|||||||||| |||
2936 TGGAG...CACAAGCAGAACTCAGTTTGTCGGGAAATAATTTAAAATATCC 2982
1724 laGluThrHisMetGln......AlaProAspSerTyrIleIleValThr 1738
     |||
2983 TGTGAAAATGACCTTCATTTATGCCGAGATAATATTTTGCCAGAGGC...... 3027
1739 CysProThrAspMetHisLeuCysAlaAspLeuThrArgHisGlyAsnPr 1755
     :::||||||||||
3028 ...ATAGATGTTCACAATGCAGAGTTCGTTCTGACTGGAGCTCACTC 3073
1755 oLysCysHisIleValSerThrGluPheValMetSerSerIleLeuArg 1772
     |||
3074 AAACCCTGGACTATGAA................TCATATAAGTTTAAC 3105
1772 lnGlnLeuGluIleGluProAsnLeuIleProTyrLeuTyrAsn 1786
     ||| |||
1739 CysProThrAspMetHisLeuCysAlaAspLeuThrArgHisGlyAsnPr 1786
```

seq_name: /SIDS5/gcqdata/geneseq/geneseq-embl/AA2001.DAT:AAU27994

seq_documentation_block:

ID   AAU27994 standard; Protein; 104 AA.
XX
AC   AAU27994;
XX
DT   18-DEC-2001 (first entry)
XX
DE   Human contig polypeptide sequence #147.
XX
KW   Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;
KW   mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukaemia;
KW   cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;
KW   nervous system disorder; inflammatory disorder; cell differentiation;
KW   angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;
KW   genetic disorder; bone regeneration; tendon; ligament; tissue repair;
KW   cytostatic; antirheumatic; antiarthritic; vulnerary; antiinflammatory;
KW   antibacterial; immunosuppressive; vasotropic; antidiabetic; antiparkinsonian;
KW   neuroprotective; osteopathic; antidiabetic; antiasthmatic; antiallergic;
KW   immunostimulant; analgesic; gene therapy.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
XX
PN   WO200164834-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US049926.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
PR   17-JUN-2000; 2000US-0597707.
PR   14-JUL-2000; 2000US-0616807.
PR   19-SEP-2000; 2000US-0664641.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;
PI   Drmanac R;
XX
DR   WPI; 2001-589862/66.
DR   N-PSDB; AAS44894.
XX
PT   Novel polypeptides and nucleic acids obtained from cDNA libraries
PT   prepared from various human tissues, for diagnosis, treatment of
PT   cancer, neurological, inflammatory disorders and for use in arrays for
PT   detection .
XX
PS   Claim 10; Page 144; 153pp; English.
XX
CC   Sequences AAU27676-AAU28019 represent full-length polypeptides and
CC   contig polypeptides of the invention. The proteins and their associated
CC   DNA sequences are useful for the treatment, diagnosis and prevention of
CC   various types of disorder in a mammalian subject such as a human, dog,
CC   monkey, mouse, hamster or rat. The disorders include cancers such as
CC   leukaemia, lymphoma and neuroblastoma, autoimmune disorders such as
CC   multiple sclerosis, connective tissue disease, rheumatoid arthritis,
CC   diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system
CC   disorders such as Parkinson's disease, Alzheimer's disease, Huntington's
CC   chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and
CC   Wernicke disease, inflammatory disorders such as nephritis, Crohn's
CC   disease, ischaemia-reperfusion injury, shock, sepsis and inflammatory
CC   bowel disease. The sequences exhibit activity relating to angiogenesis,
CC   cell proliferation, cell differentiation, stem cell growth factor,
CC   activin or inhibin. Therefore, they can be used to neuroepithelial cells
CC   in culture to give rise to neuroepithelial cells that can be used to
CC   augment or replace cells damaged by illness, accidental damage or genetic
CC   disorders. The sequences may also be used for regeneration of bone,
CC   cartilage, tendons and ligaments and in tissue repair and burn healing.
CC   Note: Some sequences for this patent did not form part of the printed
CC   specification, but were obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

SQ   Sequence   104 AA;

alignment_scores:
         Quality: Length:
         Quality: 319.00   Length: 96
         Ratio: 4.833   Gaps: 1
Percent Similarity: 68.750   Percent Identity: 68.750

alignment_block:
US-09-664-641-10 x AAU27994

Align seq 1/1   to: AAU27994   from: 1   to: 104

2833 TACATCACACCTGGAATCTCCCCAAGTCTTCCACTATGAAGGCAATCGT 2882
     ||||||||||||||||||||||||||||||||||||||||||||||||
1 TyrIleThrProGlyIleCysProSerLeuSerThrMetLysAlaIleVa 17

2883 AGAGTGTGCAGGAGGAAAAGGTGTTATCCAAGAGCCCATCTTTCCGGAAGC 2932
     ||||||||||||||||||||||||||||||||||||||||||||||||||
17 lGluCysAlaGlyGlyLysValLeuSerIleGlnProSerPheArgLysL 34

2933 TCATGGAGCAACAACCAGAACTCG................................ 2955
     ||||||||||||||||||||||
34 euMetGluHisLysGlnAsnSerValGlyArgValGlnSerAlaSerGlu 50

2955 2955
51 SerValIleAlaLeuTyrHisIleGlnArgSerValThrGluTyrLeuIl 67

2956 ..............AGTTTGTCGGAAAATAATTTTAATATCCTGTGAAAATG 2992
               ||||||||||||||||||||||||||||||||
67 ePhePheSerLeuSerLeuSerGlyIleIleIleIleLeuIleSerCysGluAsnA 84

2993 ACCTTCATTTATCCCAGAATATTTTCCAGAGCCATA 3030
     ||||||||||||||||||||||||||||||||
84 spLeuHisLeuCysArgGluTyrPheAlaArgGlyIle 96

seq_name: /SIDS5/gcgdata/geneseq/geneseq-embl/AA2001.DAT:AA005204
ID   AA005204 standard; protein; 143 AA.
XX
AC   AA005204;
XX
DT   06-NOV-2001   (first entry)
XX
DE   Human polypeptide SEQ ID NO 19096.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO20016835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI85135.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune

---

PT   disorders -
XX
PS   Claim 20; SEQ ID NO 19096; 1399pp + Sequence listing; English.
XX
CC   The invention relates to human polynucleotides (AA179941-AA193841) and
CC   the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   143 AA;

alignment_block:
US-09-664-641-10 x AA005204

Align seq 1/1   to: AA005204   from: 1   to: 143

alignment_scores:
         Quality: Length:
         Quality: 294.50   Length: 69
         Ratio: 4.750   Gaps: 1
Percent Similarity: 89.855   Percent Identity: 84.058

1471 CTACAGCAGCAGCAGCAGCCCAGCAGCAGCAGCAGCAGCACCCCGGTTTT 1520
     ::|||||||||||||||||:::||||:||:|||||||||||||||||||
1 ValGlnGlnGlnGlnProGlnGlnGlnGlnGlnGlnHisProValIle 17

1521 ACACCTTCAGCCCCAGCAGATAATGCAGCTCCAGCAGCAGCAGCAGCAGC 1570
     |||||||||||||||::||||:::||||:||:|||||||||||||:::|
17 uHisLeuGlnProGlnGlnIleMetGlnLeuGlnGlnGlnGlnGlnHisProVallLe 34

1571 AGATGTCTCAGCAACCTTACCCCAGCAGCGCCGCCATCCATTTTCACAG 1620
     ||||:||||||||||||||||||||||||||||||||||||||||||
34 InIleSerGlnGlnProTyrProGlnGlnProProHisProPheSerGln 50

1621 CAGCAGCAGCAGCAGCAACAGCAGCCACCACCA...............TCGCCTCAGCA 1661
     ||||||||||||||||||||||||||||||| ||||||
51 GlnGlnGlnGlnGlnGlnGlnGlnAlaHisProHisSGlnPheSerGlnGlnGlnGl 67

1662 GCATCAG 1668
     | ||
67 nLeuGln 69

seq_name: /SIDS5/gcgdata/geneseq/geneseq-embl/AA1999.DAT:AAY03183
ID   AAY03183 standard; protein; 1522 AA.
XX
AC   AAY03183;
XX
DT   16-JUN-1999   (first entry)
XX
DE   Topoisomerase II binding protein 2.
XX
KW   Topoisomerase II binding protein; TopBP; anticancer agent.
XX
OS   Homo sapiens.
XX
PN   JP11075856-A.
XX
PD   23-MAR-1999.
XX
PF   17-SEP-1997; 97JP-0251544.
XX

PR   17-SEP-1997; 97JP-0251544.
XX
PA   (CHUS) CHUGAI PHARM CO LTD.
PA   ('TSUR.) TSURUO T.
XX
DR   WPI; 1999-257704/22.
DR   N-PSDB; AAX28154.
XX
XX
PP   New Topoisomerase II- binding protein - useful as an anticancer
PP   agent
XX
PS   Claim 7; Page 14-18; 28pp; Japanese.
XX
CC   This sequence represents the topoisomerase II binding protein (TopBP) of
CC   the invention. The TopBP protein is useful as an anticancer agent. TopBP
CC   can be used as the target molecule for anticancer agent.
XX
SQ   Sequence 1522 AA;

alignment_scores:
         Quality: 263.50   Length: 1031
         Ratio: 0.620   Gaps: 43
         Percent Similarity: 41.222   Percent Identity: 19.011

alignment_block:
US-09-664-641-10 x AAY03183

Align seg 1/1   to: AAY03183   from: 1 to: 1522

234 AATGCACTAGCCTCACACAATAATCTCAGAGAGATGGGGACAAATCCACAGGT 283
140 AsnValSerValThrHisLeuIleAla.......... GlyGluVa 151

284 CGGGAGAAGCTCGGGAAGTC................TTTGACTTACCTGTTGTAA 321
151 lGlySerLysLysTyrLeuValAlaAlaAlaAsnLeuLysLysProIleLeuL 168

322 AGCCTCTCTTGGGTGATTCTG............................TCCGTT 347
168 euProSerTrpIleLeuThrLeuTrpGluLysSerGlnGluLysLysIle 184

348 CAGTGTGGAACTCTTCTGCCAGTAAAATGTTTTTTCTCAGAATCATGTCA 397
185 ThrArgTyrThrAspIleAsnMetGluAspPhe............LysCysPr 198

398 CATTTTTTTTTGGAATCACTCCCTTGCCTTTCTCACGGTGTTGATACAAAGCT 447
198 oIlePheLeuGlyCysIleIleCysValThr............ 208

448 GGAGCTCTTTCTTGCTGGAGTCTTTCCAAGCTCTCCAAGCAGAGGTAGGGAA 497
208 208

498 GGGAGCTTGTTCCAGCGAGAAGTTGGGAAGCACAGATCATCTGCCTTCTT 547
209 ....GlyLeuCys.GlyLeu.... 213

548 CTCACCCGGTATTCATGCAGGCTGAGGCCTCTGTTGTTGTAATGCTGGGTG 597
214 ....AspArgLysGluValGlnLeuPheThrValLysHisGlyGl 227

598 TCATCTGAAGACAGAAGTGCCCTGTGGGCCTTTGGTTACGTTCTATGGGGG 647
227 yGlnTyrMetGlyGlnLeuLeuLysMetAsn...GluCysThrHisLeuLeuV 243

648 AGATTGC.......CAGTTAACCCTCAATAAGAAAATGCACCCATTTGATTG 691
243 alGlnGluProLysGlyGlnLysSerTyrGlnLysSerTyrGluCysSerAla... 258

692 TTCCAGAGCCAAAAGGGGGAAAATACGAATGTGCTTTAAAAGCGAGCAAGCT 741

AATTAAAATTGTGACTCCTGACTGGGTTCTTGGAATCGGTATCAGAGAAAC 791
ValHisCysValThrThrArgGlnTrpPheAspPheSerIleGlyIleGlyArg 275

CAAAAAGGACGGAAGCATTTTATCATCCTCGTTCTGCAGTTATTTTATGAAGAGG 841
eCysGlnAspGluSerIleTyr... 283

AAGAAGAGGAAGAGGAAGGAGAAGGAGGAAGTAGAAAATGAGGAACAAGAT 891
ySThrArgGluProArgGlnAlaLysThrMetProAsnSerSerThrPro 299

TCTCAGAATGAGGGTAGTTACAGATAGAAGCTCAAGCCCTGCCAGCTCTCA 941
ThrSerGluIleAsnThrIleAspSerArgThr... 310

AGAGAGGGTCTCCTTCAGGTGACCAGCAGTTTTCACCTAAAACCAACACTG 991
.......LeuSerAspValSerAsnIleS 318

AAAAAATCTAAAGGGGGATTAATGTTTGATGANTTCTTCAGATTCATCACG 1041
uAspValSerAlaPheGlnAlaProGluAspLeuAlaSerGlyLysArgI 364

TTAACTTGTGTGCCAATGTCCCACCCGTCCCAGGTAACATTTTGCCCCCT 1188
leTyrLeuTyrLeuSerGly... 368

GAGGTCCGGGGGTAATTTAATGGCTGCTGGACAAAAAACCTCAAAGTTCTGA 1238
SerLysSerGluProThrLeuGlu... 347

AAGATCAGAAATGAATAGCTACCTACGGTCCAGCTGTACGGACACTGAGGA 1288
.............PheSerGlyArgGlyLeuAspLys... 376

ATATTAACTAATAATGCTGACATTCAGACAGATGAACCGGCCATCAAAATGTA 1338
snSerGlyGlyValGlyValArgPheAsnGlnLeuAsn...GluAspVal 396

GCACATATCTTACAGACTCTTTCAGCACCTACGACAAAAAATTTAGAACAGCA 1388
ThrHisValIle....ValGlyAspTyrAspsApGlyGluLeuLysGlnPh 411

GGTGAATCACAGC...CAGCAGGGACAATACAAAATCCCAATGCAGTGCTGT 1435
eTrpAsnLysSerAlaHisArgProHisSerValAlaGlyAlaLysTrpLeu 427

TTAGCCAAGTGAAAGTGACTCCAGACACACAACATCCTACAGCAGCAGCAG 1485
LeuGlyCysPheSerLysGlyTyrMetLeuSerGluGluPro 441

CAGGCCCAGCAGCAGCAGCAGCAGCCGGTTTTACACCTTCA 1529
TyrIleHisAlaAsnTyrGlnProValSerProHis...Gl 457

GCCCCAGCAGNATAANTGCAGCTCCAGCAGCAGCAGAGATCTCTC 1579
nProGluSerLysAlaLeuLeuLysLysLysAsnSerSerPheSerL 474

AGCAAACCTTACCCCACGCCACCGCCCCGCCAT.................... 1608
ysLysAspPheAlaProSerGluLysHisGluGlnAlaAspGluAspLeu 490

```
1608  ...........................................................................  1608
491   LeuSerGlnThrTyrGlnAsnGlySerThrValValGluAlaLysThrSe  507
1609  .......CCAtTTTcAcAGcAgccAGcAGccAGcAA...................  1635
507   rGluAlaArgProPheAsnAspSerThrHisAlaGluProLeuAsnAsps  524
1636  .................CAGCAGcAccAccAtcCCCTcAGcAGCCATcAG  1668
524   erThrHisIleSerLeuGlnGluGluAsnGlnSerSerValSerHisCys  540
1669  CTTTTTGGACATGATCCAGGAGTGGAG...ATTTCCAGAAGAGAGGCTTCTT  1715
541   Val.........ProAspValSerThrIleThrGluGlyGluProPh  553
1716  ATTGGGATGTCGTGTTTGcAAtTGCCGATTATCCAGACAGcAGAGTGTGATA  1765
553   eSerGlnLysSerPheLeuValGlyPheSerAsnGluAsnGlu....  568
1766  AGCCAACTGCTGCCCACCTGGAAAAGCATAATCCAGACAGCACATGGCGGCACT  1815
569   ........SerAsnIleAlaAsnIleIleLeuLysGluAsnAlaGlyLys  581
1816  GTTGACCCCACCTTCACGAGTCGATGCATGCACGCAC..................  1848
582   Ile...MetSerLeuLeuSerArgThrValAlaAlaSpTyrAlaAlaValPr  597
1849  ...CTTCTCTGAGAGTCAAGTCCAGAGCCGGCGTATGCACAGGCGGAATAA  1894
597   oLeuLeuGlyCysGluValGluGluGlnHisAlaThrValGlyIly....  608
1895  CAGAAACGAAACAGGATGTCGTTACTGCACACTGGTTAAACACAGTCTTAAAA  1944
609   ..........SerAsnIleAlaAsnIleIleLeuLysGluAsnAlaGlyLys  621
1945  AAAAAAAAATGGTA...CCGCCGCACCGAGCCCTTCACTTCCCAGTGGGC  1991
622   TyrGlnThrLeuPheAspProLysSerAsnProLeuPheThrProValPr  638
1992  CTTCCCACCAGGAGAGAAAGCCAATGTTCAGACGTATTATTTTCTGTGACTG  2041
638   oValMetThrGlyMetThrProLeuGluAspCysValIleSerPheSerG  655
2042  CAGTTGTTTTGATAGTGACACAGAGATGACCTAAAAATTAATTAATGGCTTTATTTGGCA  2091
655   lnCysAlaGlyAlaGlyLysGluSerLeuThrPheLeuAlaAsnLeuLeu  671
2092  GGTGCCAAAATATACGGGTTATCTATGCCGC.....................  2121
672   GlyAlaSerValGlnGluThrTyrPheValArgGlyLysSerLysLysGl  688
2122  .....AGCAACACGTCTCATCTGTAGAACAACAACTGCTTTAAAGT  2164
688   yMetPheAlaSerThrHisLeuIleLeuLysGlyArgGlyGlySerLysG  705
2165  ATGAAAAAGCCAAAGAGTGGAGATACCCTGTGTCAACCCCAGTGGCTT  2214
705   yrGluAlaLysAlaLysLysTrpAsnLeuProAlaValThrIleAlaTrpLeu  721
2215  .........................GGGACATTCTTCT  2228
722   LeuGluThrAlaAlaArgThrGlyLysArgAlaAspGluSerHisPheLeuIl  738
2229  GGGAAAC..............................  2235
738   eGluAsnSerThrLysGluGluArgSerLeuGluThrGluIleThrAsnG  755
2235  .......................  2235
755   lyIleAsnLeuAsnSerAspThrAlaGluHisProGlyThrArgLeuGln  771
2236  ...........................TT  2237
```

```
772   ThrHisArgLysThrValValThrProLeuAspMetAsnArgPheGlnSe  788
2238  TGAGGCACTGAGGCCAGATT......CAGTATAGTCCTCACCGGCATTCA  2281
788   rLysAlaPheArgAlaValSerGlnHisAlaArgGlnValAlaAlaAlaS  805
2282  GT....................................  2283
805   erProAlaValGlyGlnProLeuGlnLysGluProSerLeuHisLeuAsp  821
2283  ......................................  2283
822   ThrProSerGlyPheLeuSerLysAspLysLeuPheLysProSerPheAs  838
2284  ..................CTGCAGGATCCATTGCCCGTACCCAGC  2311
838   pValLysAspAlaLeuAlaLeuAlaLeuGluThrProGlyArgProSerGlnG  855
2312  AT.........................TTAGTTTTAAATCTT  2328
855   lnLysArgLysProSerThrProLeuSerGluGluValIleLeuValLysAsnLeu  871
2329  TTAGATGCTTGGAGAGTTCCCTTTAAAAGTGTCTGCAGAGTTGTTGATGAG  2378
872   ...........GlnLeuAlaLeuAsnSerArgAsnAlaValAl.  884
2379  TATAAGACTACCTCCCAAGACTGAAACAGAATGAAGTAGCTAAATGTCCAGC  2428
884   aLeuSerAlaSerProGlnLeuGlySGlu.............  893
2429  CTTCTTCCCAAAGAGCCAGAAATTGAAGACGTACACCCCCACTAAAAG  2478
894   .......AlaGlnSerGluLysGluGlnAlaProLysProLeuHisLys  907
2479  CTAACTCCAGAATTGACCCCTTTGTCCTTTTCACTGGATTCGGAGCCTGT  2528
908   ValValValCysValSer....................  913
2529  CCAGGTTCAACAGTATATTAAAGAAGCTCTACATTCTTGGTGTGGAGAGGTTG  2578
914   .......LysLysLeuSerLysLysGlnSerGluLeuGluAsnGlyIleAlaA  928
2579  CGGAGTCTGCA..........CAGAATGCACACAC  2604
928   laSerLeuGluLysGluGlnLysAlaAspIleTyrArgTrpSerPheAspGluThrValThrHis  944
2605  CTCATT.......GCCAGCAAAGTGACTCGCACCGTGAAGTTCCT  2642
945   PheIleTyrGlnLysGlyArgProAsnAspSerThrAsnArgGluThrTyrLysSerVa  961
2643  GACGGCGGATTTCTGTCGTGAAGCACATAGTGACGCCCAGAGTGGCTGGAAG  2692
961   lLysGluArgGlyVal.....HisIleValSerGluHisIleTrpLeuGlyAlaA  976
2693  AATGCCTTCAGGTGTCAGGAAGTTCAATGATGAGAGGAGAACTAC  2733
976   spCysGlyAlaGlyInGluCysLysHisLeuProGluSerLeuGlyThrArgLeuTyr  989
```

OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS87786.
XX

PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX

PS   Claim 20; SEQ ID No 53958; 103pp; English.

XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   61 AA;

XX
XX

alignment_scores:
        Quality: 249.00   Length: 84
        Ratio: 6.385   Gaps: 5
Percent Similarity: 46.429   Percent Identity: 45.238

alignment_block:
US-09-664-641-10/rev x ABG23599

Align seq 1/1   to: ABG23599   from: 1 to: 61

1650 TGGTGGTGGCTGCGTCGTTGCTCTCCTCGTCGCTGCTGCTGAAAATGGATGCGCCG 1601
     |||||||| ||||||||| ||||||||||||||||||||
   7 TrpTrpTrp.CysCysCysCysCysCysCysCys........CysCysC 20

1600 GCTGCTGCGGGGTTAAGGTTGCTGAGAGATCTGCTGCTGCTGCTGCTGCTGG 1551
     ||| ||||||||||||||||||||||||
  20 ysCys..............CysCysCysCysCysCysCysCysCysCys 32

1550 AGCTGCATTATTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGCTGCTG 1501
     ||| |||||||||||||||||||||
  33 CysCysCysCysCys..........CysCysCysCysCysCysCys 45

1500 CTGCTGCTGCTGGGGCTCGCTGCTTGCTCCTTGTAGCATGTGTGTCTGTCTCGAGTCA 1451

---

seq_name: /SIDS5/gcgdata/geneseqp-embl/AAA2001.DAT:ABG19008
seq_documentation_block:
ID   ABG19008 standard; Protein; 398 AA.
XX
AC   ABG19008;
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
DE   Novel human diagnostic protein #18999.
XX
DT   18-FEB-2002 (first entry)
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS82495.
XX

PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX

PS   Claim 20; SEQ ID No 49367; 103pp; English.

XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   398 AA;

XX
XX

alignment_scores:
        Quality: 248.00   Length: 157

1450 CTTTCACTTGGCTAAAACACACTGCATTGGCATTTGTAGTGTCCCTGCTGG 1401
     | |||||| :||||||||||||||| ||||||| |||
  45 sCysCysCysCysCysCysCysCysCysCysCysCysCys............ 57

  58 ....CysCysCysTrp 61

Percent Similarity: 40.764   Percent Identity: 29.936

Ratio: 3.875   Gaps: 7

alignment_block:
US-09-664-641-10/rev x ABG19008

Align seq 1/1 to: ABG19008 from: 1 to: 398

1730 AACACACATCCCAATAAGAAGCCGTTCTTCTCGAATCCCACTGCT..... 1686
         :::::||| |||||||| :|:||||:|||:||
138 SerThrSerAspSerGlnSerProCysGlyValGlySerLeuVa 154

1685 ....G 1685
                                                           |
154 lVal***AlaGluArgSerCysAlaGlnCysArgMetSerProArgLysC 171

1684 GATCCATGTCCAAAAAACCTGATGCCTGCTGAGCGGATGCGTGGCGTGCTGT 1635
        ||| ||||||
171 lysIlsCys..........................CysCys 177

1634 TGCTGCTGCTTGCTGCTGTGAAAATGGATCGCGGGGGGGTTGCTGGGGGTAAGG 1585
        ||||||||||||||||| ||| ||
178 CysCysCysGlyCysCysCysSerCysCys......CysTyrCysHisTy 192

1584 TTGCTGAGAGATCTGCTGCTGCTGCTGCTGGAGCTGCATTATCTGCT 1535
        ||||||||||||||||||||||||||||||| ||
192 rCysCysCysTyrCysCysCysCysCysCysCysCysCys...... 208

1534 GGGGCTGCAAGGTGTAAAACCGGTGCTGCTGCTGCTTGCTGCTGGGGCTGC 1485
        ||||||||||||| |||
209 ....CysCysCysCysCysCysCysCysCysCys 221

1484 TGCTTGCGTGCTCTAGCCATGTGCTGTCTCTCGGAGCTGCACTTTCACTCGGCTAAA 1435
        |||||||||| |||
222 CysCysCysCysTyrCysCys.......... 228

1434 CAGACACTGCATTGGCATTGTATGTCCCTGCTGGCGTGTGATTCACCTGCT 1385
        ||||||||||| :::
229 ....CysArgCys..........TyrCysCysC 235

1384 CTTCTAAATTTTTTCCGTTAGTGTCGCTGAAAGAGTCGTAAGATATGTGCTACA 1335
        ||||| :::
235 ysCysCysTyrCysCysCysCysCysAsArgCysCysCysCysCys 251

1334 TTTCATCGCCGGTTCATCTCC 1314
        |||
252 CysCysCysArgCysCysCys 258

seq_name: /SIDS5/ocgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG23600

seq_documentation_block:
ID   ABG23600 standard; protein; 87 AA.
XX
AC   ABG23600;
XX
DT   18-FEB-2002 (first entry)
XX
DE   Novel human diagnostic protein #23591.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-693362/73.
DR   N-PSDB; AAS87787.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity –
XX
PS   Claim 20; SEQ ID NO 53959; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantifying (II), and its binding partners, as molecular weight markers and as
CC   a food supplement. (II) and (II) are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   87 AA;

alignment_scores:
                        Quality: 242.00   length: 127
                        Ratio: 4.566   Gaps: 7
Percent Similarity: 41.732   Percent Identity: 36.220

alignment_block:
US-09-664-641-10/rev x ABG23600

Align seq 1/1 to: ABG23600 from: 1 to: 87

1742 TCGGCAATTGCAAACACATCCCAATAAAGAAGAGCTTCTTCTGGAATCTC 1693
        ||| |||||||::: ||||:|||:|:||:||
5   SerLeuIleAlaSer.......ProArgSerLeuProSerAlaProGlu 19

1692 CACTGCTGA......................TCATGTCCA.. AAAAGCTGAT 1664
        :::|||||||| ||||||| :|||| |
19 uValAlaGlyLeuSerHisGlyCysGluLeuCysProAspArgSerProc 36

1663 GCTGCTGAGCGGATGGTTGGTGGCTGCCTGCTGCTGCTGCTGCTGCTGAA 1614
        |||||||||||||||||||||||||||||||||||||||||||
36 ysCys.............CysCysCysCysCysCysCysCysCys.. 45

1613 AATGGATGCGGGCGGGGCTGCTGGGGGTAAGGTTGCTGAGAGATTGCTGCTG 1564
        ||||||||||||||||||||||||||||||||||
46 ....CysCysCy 48

1563 CTGCTGCTGCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCG 1514
        |||||||||| |||
48 sCysCysCysCysCysCysCysCysCys............ 57

1513 GGTGCTGCTTGCTGCTTGCTGCGCCCTTGCTGCTGCTGCTGCTGTAGCCATGTGT 1464
        ||||||||||||||| :::|||||||||||||||||||| |||
58 .CysCysCysCysCysCysCysCysCysCysCysCysCysCysCysCys 73

```
1463 GTCTCTGGAGTCACTTCACTTGGGTAAACAGCACTGCATTGGCATTTGT 1414
   74 .................................................Cy 74

1413 ATGTCCCTCGCTCGGCTGTGATTCACCTGCTGT 1383
     ||||  |||   |||   ||||  |||||
   74 sCysCysCysCysCysPheCysCys 84
```

seq_name: /SIDS5/gcgdata/geneseqp-embl/AA2001.DAT:AAU07343

alignment_block:
US-09-664-641-10/rev x AAU07343   ..

alignment_scores:
       Quality: 238.00
         Ratio: 2.833   Length: 193
    Percent Similarity: 43.523   Gaps: 10
                                   Percent Identity: 30.052

Align seq 1/1 to: AAU07343 from: 1 to: 1679

```
1823 GGGTCAACAGTGCCGCCCATGTGCCCTGGATTATCCTTTTTCCAGGTGGCCAG 1774
     ||:||:|||||||:::       :||||||::|||||||||||||||||||
   25 GlyAlaThrThrCysGlyCysCys.................................32

1773 CAGTTGCTTATCAGACATCTCCTTGGATAATCCGAATTGCAAACACAC 1724
     ::||||      ||      :::   |||  |||||||||||||||:
   33 .GlyCysGlyThrCysCysGlyThrGly.........................41

1723 ATCCCAATAAGCAAGCCTTCTTCTGGAAATCCACTGCTGGATCATGTCCA 1674
                     ||||||:|||:::|||:|:|||
   42 ........................CysThrGlyGlyAlaCysGly 48

1673 AAAAGCTGCATGCTGCTGAGGCCGCCATGGTGGTGGCTGC................1638
     ||||||||      |:|||      ::|||||||||||
   49 ..............CysCysThrGlyAlaGlyAlaGlyCysGlyCysGly 62

1637 TGTTGCC...TGCTGCTGCTGCTGTGTGAAAATGGAT 1607
     |||||       ||||||||||||||||||||:|
   62 aCysCysCysCysGlyCysCysGlyCysCysGlyGlyCysCysThrCysCysAlaC 79

1606 GGGGGCTGCTGGGGGGTAAGGTTGCTGAGAGATCTGC 1569
     :::|||       |||||||||||||||||||
   79 ysCysCysCysAlaGlyGlyCysThrGlyGlyCysThrCysGlyGly 95

1568 TGCTGCTGCTGCTGCTGAGCCTGCAATTATCTGCTGGGGCTGAAGGTGTAA 1519
     |||||||||||       :::||||||||||||||||||       |||
   96 AlaCysCysCysCysGlyCysGlyCysGlyGlyCysCysCysSerCysCysAl 112

1518 A.........ACCGGGTGCCTGCTGCTGCTGCTGCTGGGGCCCTGCGTGCTGCT 1478
     ||||||||||||||||||||||||||||||||||||
  112 aCysThrGlyThrCysGlyCysCysGlyCysCysCysCysThrCysCysThrC 129

1477 CCTGTAGCACATGCTGTCGTCTGGAGTCACTTCACT................1443
     ||       |||       :::|||||       |||
  129 ysGlyGlyAlaCysCysThrGlyThrCysGlyGly 145

1442 ..................TGGCTAAACACGACTGCAATTGGCCATT 1417
     :::||||||||||||||||:::      ::
  146 GlyAlaGlyAlaGlyCysThrGlyGlyAlaGlyAlaThrGlyThrThrCy 162

1416 TGTATGTCCCTGCTGCTGCTGATTCACCTGCTGTTCTAAAATTTTCGTAG 1367
     ::||||||||||||||||||||       :::|||||||||
  162 sAlaCysGlyCysCys........ThrThrCysCysThr...CysAlaThr 175

1366 GTGCTGAAAGAGTCTGTAAGGATATGCTGCT 1338
     :::|||       ||||||
  175 laAlaGlyGlyAlaCysThrThrCysAla 184
```

seq_name: /SIDS5/gcgdata/geneseqp-embl/AA2001.DAT:ABB69280

```
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13383.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 34632; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL6176-ABL30511), expressed DNA
CC   sequences (AHL0184O-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   776 AA;
XX

alignment_scores:
         Quality:    225.50                    Length:        859
         Ratio:      0.616                      Gaps:           40
Percent Similarity:  42.608         Percent Identity:   21.886

alignment_block:
US-09-664-641-10 x ABB69280

Align seg 1/1  to: ABB69280   from: 1   to: 776

778 GTATCAGAAAAACCAAAAAAGGACACCAACCAYTTYATCACCTCGTCTCTGAT  827
    ::::::||||||    ||||||   |||||||||||||||||||||||||:::
 15 LeuAlaGlnGlnLysLysLeuProLeuAlaTyrGlnThrAsnLeuVa  31

828 TATTTATGAAGAGGAGGAACAAGAGGAAGAGGAAGAGGAAGAGGAAGTAGAAA  877
    : |||:||                                       ||||:
 31 lAspTyrArgThrAla........................    36

878 ATGAGGAACAAGATTCTCAGAAATGAGGGTAGTACAGATGAGAAGTCAAGC  927
    :::||||::    ::::::::                 |||     |||
 37 ....AlaTyrSerGlnAlaLeuTyrGlnGlnSerProThr  48

928 CCTCCCAGCGTCTYCAAAGAAGGGTCCCTTCAGTGACCAGCAGTTTCACC  977
    ::||||||    :::::::::||||||||||||||||||||||||||
 49 AlaThrSerSerGlyGlyGlyProLeuSerProIleAspMetGlnPr  65

978 TAAATCC..............AACACTGAAAAATCTAAAGCGGAATTAATAGTTTG  1018
    :::||||            |||           ::::|||       :
 65 oGlnSerLysHisTyrAsnLeuMetLysHisGlyGlyThr........S  80

1019 ATGATTCTCAGATTCATTCACCGGAAAAACAGGAG..............  1053
     |||||||||||    ||||||||||                        
 80 erSerSerHisSerSerProTyrHisGlnSerTyrSerSerSerGly  96

1054 ..............ACAAATTTAAACTGACACCCCGCCCGCAAGTCCCACA  1088
              |||||||||                     ||||||||
 97 GlyThrAlaAlaGlyvluGluLeuTyrGlnSerProThrGluArgThrTy  113

1089 GTTAGCTGCACCGAAAAACGAGGGTCCCTCACGAGAAAGAGCTCGGGTTGA  1138
     ||||||||||||||       |||                    |||
 113 rLeuAlaAlaAlaGlyArgLeuHisAlaAsn.............  123

1139 TTAACTTGTGTCGCCAATGTCCCACCCGTCCCAGGTTAACATTTTGCCCCCT  1188
     ::||||||||||||||||                        :::|||||
 124 ..AsnAlaMetAlaGlyHisHisProIleSerAla.  134
```

```
1189 CAGGTCCCGCGGGTAAATTTAAATTGGCTCGTCGGACAAAACCTCCAAAGTTCTGA  1238
     ||||||||||||||||||||||||||                :::||||||||:::
 135 .............LeuGlnSerGlnTyrGlnGlnLeuGlnAlaAlaAlaLy  146

1239 AAGATCAGAAAATGAATAGCTACCTACCTGGAGTCCCAGCCTGTACGG..........  1278
     :: :::||||||:||||| |||| |||   |||      :::||||||
 146 sMetGlnAlaGlnMetAlaThrGlnSerGluAlaAlaAlaGlnGlnGlnA  163

1279 ..........ACACTGAGGAAATATTACTAATAAATGCTGACAATTCAGCAGATG  1320
           |||||||||||||||||                |||     |||
 163 rgThrPheAlaLeuArg.........................GlnAlaMet.  171

1321 AACCGGCCATCAAAATGTAGCACATATATCTTACAG..............  1353
     |||  ||||||||:||||||||||||||||||||||             
 172 AsnProProThrAsnHisTyrHisMetSerGlnSerSerSerMetValSe  188

1354 .ACTCTTTCAGCACCTACGAAAAATTTAGAACAGCAGGTGAATCACAGCC  1402
     :||||||||||:||||::::||||          :::::  |||||||||
 188 rThrPheMetThrAlaLeuAsnGlnGlnGlnGlnGlnGlnGln.......  200

1403 AGCAGGGAGCATACAAATGCCAAATGCACTGCTGTTTAGCCAAGTGAAAGTG  1452
     |||||||||||    ||||||||||||||||||||||||||||||||||||
 200 ...............GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln  200

1453 ACTCCAGAGACACACATGCTACAGCAGCAGCAGCAGGCCCAGCAGCAGCA  1502
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 200 ...............GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGl  210

1503 GCAGCAGCACCCGGGTTTTACACCTT..............  1534
     ||||||||||||||||||||||||||            ||
 210 ....SerGlnGlnProAlaMetProLysHisTyrGlnGlnGlnLeuTyrAlaGl  227

1535 AGCAG...........ATAATGCAGCTCCAGCAGCAGCAGCAGCAGAGATC  1575
     ||                ::||||||||||||||||||||||||||||||
 227 laGlnGlyProLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln  243

1576 TCTCAGCAACCT..............  1589
     ||||||||||||            ||
 244 SerGlnHisGlnProAlaMetProLysHisTyrGlnGluGlnLeuTyrAlaGl  260

1590 CCCCCAGCAGCCGCGCCGCCATCCATTTTCACAGCAGCAGCAGCAGCAACAGC  1639
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 260 nGlnGlnLeuGlnLeuGlnLeuHisGlnGlnGlnGlnGlnGlnGlnGlnGl  277

1640 AGCCACCACCATCGCCTCAGCAGCATCAGCGTTTGGACATGATCAGCA  1689
     |||     ::::|||||||||||||||||||:::|||||||||||||||
 277 lnLeuGlnLeuGlnGlnGlnGlnHisArg.........HisLysAsnAsp  290

1690 GTGGACAGTTCCAGAAGAAGGCTTCTTATTCGGATGTGTTGTTTGCAATTGC  1739
     :::::                ||||||||||||||||||||||||||||
 291 LeuGlnThrPro.....GlySerGluHisGlyThrValTyrIleGlnGl  305

1740 GGATTATCCAGGAGCCAGAATGTGTTAAGCAAACTGCTGGCCACCTGGAAAA  1789
     :::|||||||:||||||||||::::||||||||||||||||||||||||
 305 nAsnHisProAsnHisValValAsnGln........................  314

1790 GGATAATCCAGGGACACATGGGCGCCACTGTTGACCCCACCTTCACGAGTCGA  1839
     |||||                      ||||||||||||||||||||||||
 314 ..............................AlaThrValAspProThrPheThrSerAr  327

1840 TGCACGCACCTTCTCTGCTGGAGTCAAGTCACGAGCCGCGTTATGCACAGC  1889
     ||||||||||||||||||     :::||||||||||||||||||||||||
 315 .............AlacysGlnThrGlnIleSerAlaValLysProLysSerAl  327

1890 AATA.............AGAGGAAGAGAGATGTGTTACTGCACACTGGTTAA  1930
     ||||           ::|||   |||||||||||||||||||||||||:::
 327 aThrProSerGluSerSerSerAsnSerThrLysSerProSerH  344

1931 ACACAGTCTTTAAAAAAAAAAAAAAATGGTACCGCCGCCGCACGAGCCCTTCAC  1980
     ::|||   ||||||            :::||||||||||||||||||||
 344 ..........................................  344
```

seq_name: /SIDS5/gcgdata/geneseq/geneseq-embl/NA2001.DAT:ABB63516

```
344  isAlaProLeuAspArgLysLysSerAlaGlySerIleGlnAlaLeuLys 360

1981 TTCCCAGTGGCGCTTCCCACCAGGAGGAAACCCATGTTCACGCATATTAT 2030
     ||::|::||   :|::||||||:||||||:|||||||||:||::
361  SerProIleThrLysArgProProSerThrPro............Va 372

2031 TTCTGTGACTGGATTTGTTGATAGTGAC...AGAGATGACCTAAAATTAA 2077
     ||||                         |||  ||||||||:::
372  lThrLeuSerGlyTrpLeuHisIleLysGlnGlySerAspGlyLeuVal. 388

2078 TGG.......................... 2080
     |||
389  TrpAlaArgValGlnTrpPheValLeuAlaLeuGlyTyrCysLeuTyrTyrIleTyrIle 405

2081 ............................CTTATTCGCAGGTG....CCAAATATAC 2105
                                 ||:||||   ::|||
405  sGlyProGluGluGlyGlyLysLeuLeuGlySerValLeuLeuProSerTyrA 422

2106 GGGGTT............................ 2110
     |||||
422  rgValSerAlaCysLeuProGluAspLysleTyrArgLysPheAlaPhe 438

2111 ............ATCTATCCCAGCACAACAGTCCTC 2136
                ||::|||::|||||
439  LysCysGlnHisIleGlnAsnMetArgThrTyrThrTrpLeuHisAlaLeuAlaAspAsnSe 455

2137 ATCTGTAAAAGAGAACCAAACTGCTTTAAAGT 2168
     :|||||                         ::|   AATGA
455  nIleSer............GlyGluAsnSerAspSer 472

2169 AAAAGCCAAACAGAGTGGAGGATTACCCTGTGTCACAGCCCAGTGGGCTTGGCG 2218
     ::|:|::|:|||||
472  lnAlaProSerGlyGluSerGluProSerValAsnSerSerLeu.As 488

2219 ACATTCTTCTCGCAAAACTTTGAGGCGACACTGCACGAGATTCAGTATAGTGCC 2268
     :||||||                         |||
488  nIleSer............GlyGluAsnSerAspSer 498

2269 TACACGCGCATTCACTCT...GCAGGATCCATTGCCCCCTACCCAGCAGTTT 2315
     :||||  ::|||||||  ::|||||                 |||
498  leHisThrLeuGlnSerHisProGlyIleLysGlyGlnProThrProSer... 513

2316 AGTTTTAAAACTTTTAGATGCTTGGAGACTTCCCTTAAAAGTGTCTGGCAG 2365
     ::|  |:::|||||||
514  ......SerGluAsnThrGlySerSerGlyGlyGlySer.GlySerG 527

2366 AGTTGTTGATGAGTATAAGACTACCTCCCAAACTGAAACAG... 2412
     ::|:||||||                          AATGAA
527  lyGlnProLeuTyrAlaAsnAlaProProLysProGlyArgArgIleAsnAsp 543

2413 GTAGCTAATGTCCAGCCTTCTTCCCAAAGACCACAGAATTGAAcACGTACC 2462
     ::||  :||||||  :::|||||
544  GlyGlyTyrTyrSerSerProSerProGlyLeuHisAsnGluGlnGlnGlnHi 560

2463 ACCTCCCACTAAAAAGCTTAACTCCAGAATTGACCCCTTTGTGCTTTTCA 2512
     :|:||||                      |||
560  sProSerArgArgValGluMetSerPro................... 568

2513 CTTGGAATTCGCAGCCTGTCCAGCTTCAACAGTATATTAAGAAGCTCTACATT 2562
     ::|||||                       |||
569  ..............ThrGlnIleLeuTyrGlnGlnGlnGlnHisGln 579

2563 CTTCGTCGAGAGGTTGCGGCAGGTTGCACAGAGTCCACACACCCTCAATTGC 2612
     ::||                        |||
580  ArgSerGlnGlnGlnGlnGlnLeuProGlnGlnGlnHisHisAlaIleTy 596

2613 CAGCAAAAGTGACTCCCACCGTGAAGTTCCTGACGCGGCGAATTCTCTGGTGA 2662
     ::||                        |||
596  rAsp......ThrArgThr............ 601

2663 AGCACACAAGTGACGCCAGAGTGGCTGGAAGAATGCTTCAGGTGTCAGAAG 2712
     |||:|   |||                   |||......G
601  lyHisValSerThrAlaLeuGlnLeuGlnGlnAla............ 612
```

```
2713 TTCATTGATGAGCAGAAGTACATCTTCCGAGATGTGAGGCAGAAGTA.. 2760
     ::||||||:|||  |||               |||||||:::
613  ......GlnGlnIleThrSerLeuAspHisLeuGlnIleGlnPheGl 626

2761 .CTTTTCTCTTTCAGCCTTGGAAGAATCCTTAAAACGGGGCACAGTTTTCTC 2809
     |||  :|||||||||:||||||                  |||
626  nGlnGlnIleLeuAspMetGluGluGlnIleIleAlaArg.......... 638

2810 CACTCTTAAGGCCAAAATATTTTACATCACACCTGGAATCTCCCAAGT 2859
     |||
638  .............................. 638

2860 CTTTCCACTATGAAGGCCAATCGTAGTGTGCAGGAGGAAGTGTTATC 2909
     ||:|||                      ||| ::|
639  LeuGlnLeuGlnAlaGlnAlaGlnGlnIleTyrGly.............Gl 651

2910 CAAGCAGCCATCTCTTCCGGAGAGTCATGGACACAGGACAGAACTCGAGTT 2959
     ::|  |||:||||||::|||||                  |||
651  uArgGlnMetTyrMetAlaLysLeuMetGlnGlnArgGlnGlnLysGlyAsn. 667

2960 TGTCGGAAATAATTTTAATATCCTGTGAAAAATGACCTTCATTTATGCCGA 3009
     |||                        |||
667  .............TyrProThrGlnGlnGlnGlnIleLeuLeuGlnAlaGl 680

3010 GAATATTTTGGCCAGAGGCA 3028
     :||  :|||
680  uArgArgThrProAspAla 686
```

CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1186 AA;

alignment_scores:
      Quality: 220.50   Length: 650
      Ratio: 0.817   Gaps: 26
Percent Similarity: 41.538   Percent Identity: 21.538

alignment_block:
US-09-664-641-10 x ABB63516

Align seg 1/1 to: ABB63516 from: 1 to: 1186

865   GAGGAAGTAGAGAAAATGAGGAACAGAGATTCTCAGAATGCAGGTAGTACAGA   914
      ||||:::|||||| :|||::::|||
449   GluGluValGlnLysGln............LeuIleGluLeuProGl   463

915   TCGAGAGTCAACCCCTTCCAGCTCTCAAGAACGGTCTCCTTCAGGTCACC   964
      :::: |||||||:::|||:::||::|| :|||||
463   yAspLeuAsnThrProValThrThrSerSerAspIleProSer.   477

965   AGCAGTTTTCACCTAATCCAACACTGAAAAATCT.......AAAGGGGAA   1008
      ||||:|||| ||| |||||:| :|||||
478   .PheProGlyProThrThrValValGlnProProIleThrGlySer   493

1009 ..............TCTTCAGATTCATCACCGGGAAAAACAGGAGAGAAAATTAAACT   1066
      :||:||||||||:|||||||||::::
494   IleGluSerGluHisCysSerProLeuGluGlnValIleSerValAlaSe   510

1067 GGACCCCGGCCGAAGTCCCACAGTTAGCTGCAGCAAAACGGCAGGCTGCCT   1116
      ||| ||||||| ||||| :|||||
510   rProValLeuSerHisSerProLeuGlyGluArgSer.   524

1117 CAGGGAAAGCAGCCTGGGTTGATTAACTTCTGTGTCCCAATGTCCCACCGT   1166
      |||||| ||||||||:::::||| |||
525   .ThrPro.   ProAlaLeuAlaIleVal.   532

1167 CCCAGGTTAACATTTTCCCCCTGAGGTCTCCGGGTTAATTTAATGGCTGCTG   1216
      |||:::||::|||:|| :||||||
533   .LysGluGluSerSerAsnSerCys.   541

1217 GACAAAAACTCCAAAGTTCTGAAAGATCAGAAATGATAGCTACCTGGAGT   1266
      ||:::||||||| ||| |||
542   AsnMetTyrProGlnHisAsnAsnAsnAsn.   551

1267 CCACTGTACGGACACTGAGGAATATTACTTTAATAATGCTCGACATTCAGCA   1316
      :|||::||||| |||
552   AsnAsnAsnAsnAsnThrThrSerSer   560

1317 GATGAACCGCGGCCATCAAAATGTAAGCACATATCTTACAGACTCTTCAGCAC   1366
      |||:::||||| :||:::||
561   SerThrThrThrThrSerLysGlnThrThrSerGlnThrGlnSer   577

1367 CTACAGAAAAATTTAGAACAGCAGGTGAATCACAGCCAGGCAGACTACA   1416
      |||:::|||:::||||||||:::|||||||
577   sValGlySerProGlyAsnHisThrGlnSerHisGln.   589

1417 AATGCCAAATGCAGTGCTGTTTAGCCAAGTGCACTCCAGAGACACA   1466
      ||||||:::||||||||||||:::::::|||
590   ....GlnGlnGlnGlnLeuGlnHisAsnAsnThrSerSerSer   603

1467 CATGCTACAGCACCAGCAGCAGGGCCCAGCAGCGCAG..........   1503
      ||||||||||||||||||||||||
604   AsnSerAsnCys.   609

609   sSerHisGlu.GlnGlnGlnGlnGlnGlnHisMetSerSerPr   626
      ||:::|||||| :::|||||||||:::
1504 ..........CAGCAGCACCCGGTTTTACACCTTCAGCCCCAG....   1536

626   rGlnGlnGlnTyrGlnHisGlnHisGlnHisIleGluHisGlnGlnLeu   642
      ||||||||| :||||||::|||||||
1537 ....CAGAT   1541

643   ProAsnHisHisHisHisHisHisHisHisHisHisSerGlnLeu   659
      :|| |||:
1542 AATGCAGCTCCAGCAGCAGCAGCAGCAGATCTCAGCAACCTTACC   1591

659   uGlnGlnLeuGlnHisGlnHisGlnHisGlnHisGlnHisG   676
      |||||||||||||||||||||:::|||||||||::|||
1592 CCCAGCAGCGCGCGCCATCCATTTCACAGCAGCAGCAGCAACAGCAG   1641

676   lnGlnHisProLeuHisGlnHisGlnHisGlnHisGlnHisGln   692
      ||||||:::|||||||||||||||||||:::|||||||
1642 CCACCACCATCGCCTCACCAGCCACGCATCAGCTTTTTGGCATGATCAGCAGT   1691

693   ....GlnHisGlnHisGlnTyrGlnHisGlnHisAlaIl   704
      :::|||||||||||||||||||||||||||:::
1692 GGAGATTCCAGAAGAGAAGGCTTCTTATTGGGATGTGTGTTGCAATTGCGG   1741

704   eGln.   705
      :::
1742 ATTATCCAGAGACAGAGTGTTGATAAGCAACTG...CTGGCCACCTGGAAA   1788

706   ....HisGlnSerGlnGlnAlaSerHisArg   719
      |||:::|||||||:::||||::|
1789 ACGGATAAATCCACAGCCCGGCCACAGTTGACCCCACCTTCACGAGTCG   1838

720   GlyIlePheThrThrThrGlyAsnAlaMetAsnAlaAlaAlaAlaAlaAl   736
      :::|||||||||||||||||||||||||
1839 ATGCACGACCACCTTCTTCGTGAGGTCAGCTACGACGGTATGCCAGG   1888

736   aAlaAlaAlaAlaGlnGlnHisGlnHisGlnHisGlnHisG   753
      ||||||||::: |||
1889 CAATAAGAAGAAGAAGAAGATGTTACTGCACACTGGTTAAACACAGTC   1938

753   lnLeuProSerProGlnLeuGlyValLeuAla.   763
      ||| |||
1939 TTAAAAAAAAAAAAATTGGTTACCCGCCCCACCGAGCCCTTCACTTCCCAGT   1988

764   ....GlyProMetSerProSerAsnSerLeuGlyAsnSerTr   777
      :::|||||||||:::
1989 GGCCTTCCCACCAGGAGGAAAAGCCATGTTCACAG.   2022

777   pGlyLeuProSerProAspLysThrMetPheGlnProProLeuPheSerT   794
      |||:::||| |||
2023 .......CATATTATTTCTGTG.   2037

794   euProAlaHisTyrAlaThrMetGlnGlnGlnGlnGlnGlnGlnGln   810
      |||:::|||||||
2038 ..............ACTGGATTGTTGATAGTGA   2057

811   GlnGlnGlnGlnGlnGlnGlnAlaAlaGlyAlaAlaProSerPr   827
      |||||||||||||:::||||||| :::
2058 CAGAGATGACCTAAAATTAATGGCTAPTTGGCAGT......GCCAAAT   2101

827   oTyrAspAspGlyArgAlaAlaAlaAlaAlaAlaGlnHisAlaGlu   844
      ||||||| |||:::|||
2102 ATACGGGGTTATCTATGCCGCAGCAAACACAGTCCCATCTGTAAAGAACCA   2151

844   euLeuGly. .LeuThrMetAspCysThrProLeuLeuLysGlnPro   859
      ::: |||||||:::|||||||:::|||
2152 ACTGGTTTAAAGTATGAAAAAAGGCCAAAGAGTGGGGAAGGATACCCTGTGTCAA   2201

860   ProPro....SerTyrAlaGlyAlaSer.   869
      |||:::   :|||
2202 CGGCCCAGTGGCTTGGGCACAATTCTCTGGGAAAACTTTGAGGGACTGAGGC   2251

869   yPheProGlyLeuGlyLeuGlyAspLeuHisSerSerHisGlnGlnLeuGlnGln   886
      ||||::: |||

2252 AGATTCAGTATTACTTCGGCTACACGCGCCATTCAGTCTTCAGCTCTCCAGGATCCATTTGCC 2301
     ||| |||||||
886 lnGlnGlnTyr 889

2302 CCTACCCAGCATTTAGTTTTAAAYCTTTTTAGATGCTTTGGGAGAGTTCCCTT 2351
     889

2352 AAAAGTGTCTGCAGAGTTGTTGATGAGTATATAAAGACTACCTCCCAAACTG 2400
     :::||| |||||||
890 .... ValArgSerGlnProLysTyrG 897

2401 .... AAACAGAATGAAGTAGCT 2418
                                          ::|||||||||:::
897 lnTrpLeuAspSerProAlaAspTyrAlaGlnGlyGlnGlyGlnValGln 913

2419 AATGTCCAGCCTTCTTCCCAAACAGAATTGAAGACGGTACCACCTCC 2468
     :::|||||| :::||:||::::
914 GlnValGlnGlnGlnGlnGlnThrLeuValLeuProGlyPt 930

2469 CNCTAAAAAAGCTAACT 2484
930 oThrSerSerAlaSerSerSerAsnAlaAlaleuGlyValLeuIleProL 947

2485 .... CCAGAATYGACCCCTTTTGCTTTCACTGGATTC 2520
                  ||||::::|||
947 ysGlnGluAsnTyrProAspMetGlnProSerSerAsnGlyThrGlyTyr 963

seq_name: /SIDS5/gcgdata/geneseqp-embl/AA2001.DAT:ABG27446

alignment_scores:
        Quality: 220.00   Length: 101
        Ratio: 5.366   Gaps: 5
Percent Similarity: 40.594   Percent Identity: 35.644

alignment_block:
US-09-664-641-10/rev x ABG27446

Align seq 1/1 to: ABG27446   from: 1 to: 89

1700 GGAATTCCCACTGCTGGATCATGTCCAAAAAAGCTGATGCTGCTGAGCGCA 1651
     ||| ||||||||||||||||| :: || |||| |||||||||
24   GlyCysSerValAlaAlaProAlaAlaAlaAlaThrCysCys 37

1650 TGGGGTGGGCCTGCTGCTTGCTGCTGCTGCTGCTGTGAAAATGGATGCGGGCG 1601
     |||||||||||||||||||||||||||||||||||||||||||
38   .... CysCysCysCysCysCysCysCysCys 45

1600 GCTGCTGGGGTAAGGTTGCTGAGAGATCTGCTGCTGCTGTGCTGTGCTGTGG 1551
     ||||||||||||||||||||||||||||||||
46   .... CysCysCysCysCysCysCys 51

1550 AGCTGCCATTATCTCGTGGGGGTCGGAAGGTGTAAAACCGGGTGCTGCTGCTG 1501
     ||| ||||||||||||
52   ...CysCysCys .... CysCysCysCy 59

1500 CTGCCGCTGCTGGGGCCTGCTCCTGCTGCTGTGTGTAGCATGTGTGTCTCTGGAGTCA 1451
     ||||||| :::|||||||||||||||||||| |||
59   scCysCysCysCysCysCysCysCysCysCysCysCysCys 71

1450 CTTTCACTTGGCTAAAACAGCACTGCATTGGCATTGTATGTCCCTGCTGG 1401
     ||||||||||||||||||||||||||||||
72   .... CysTrp 73

1400 CTG 1398
     |||
74   Leu 74

seq_name: /SIDS5/gcgdata/geneseqp-embl/AA1999.DAT:AAY60558

XX
PN   DE19818620-A1.
XX
PD   28-OCT-1999.
XX
PF   21-APR-1998;   98DE-1018620.
XX
PR   21-APR-1998;   98DE-1018620.
XX
PA   (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI   Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX
DR   WPI; 1999-602416/52.
DR   N-PSDB; AA242235.
XX
PT   New polypeptides and their nucleic acids, useful for treatment of
PT   bladder tumour and identification of therapeutic agents -
XX
PS   Claim 23; Page 338; 366pp; German.
XX
CC   This invention describes novel polypeptide fragment sequences (I) and
CC   their encoding nucleic acids (II) which are highly expressed in normal
CC   bladder tissue and have cytostatic activity. (II) are used for
CC   recombinant expression of (I) and to isolate complete genes. (I) are
CC   used to identify agents suitable for the treatment of bladder tumours,
CC   to directly treat this form of cancer (including expression from gene
CC   therapy vectors), or are used in a preparation for cancer treatment. (I)
CC   is also used for the generation of specific antibodies. (II) are
CC   identified by assembling ESTs (expressed sequence tags) from a
CC   particular tissue type before comparison of expression patterns. This
CC   allows a significantly longer fragment of the gene to be revealed, and
CC   therefore reduces the number of failures because of ESTs from different
CC   libraries representing different parts of the same unknown gene
CC   distorting the estimated frequency of occurrence in a particular gene.
CC   AAY60329-Y60591 represent protein fragments encoded by the human normal
CC   bladder tissue cDNA library derived EST fragments represented in
CC   AA242122-Z42248.
XX
SQ   Sequence   169 AA;

alignment_scores:
         Quality: 216.00   Length: 86
         Ratio: 4.800   Gaps: 7
Percent Similarity: 52.326   Percent Identity: 43.023

alignment_block:
US-09-664-641-10/rev x AAY60558

Align seq 1/1 to: AAY60558   from: 1   to: 169

1643 GGCTGCTGTTGCTGTGCTGC...............................TGCTGCTGTGAAAATGG 1609
     ||||||||| :::::||||
1608 ATGCGGCGGC....TGCTGGGGGTAAGGTGCTGAGAGA 1574
58   GlyCysCysGlyCysGlyGlyProLeuThrProTrpGlnArgAl 74

1608 ATGCGGCGGC............TGCTGGGGGTAAGGTGCTGAGAGA 1574
     ||||||||| ::|||||
1573 TCTGCTCGTGC...`TGCCTCCTGCTGGAGCTGCATTATC.......SerCysSerAsnC 89
74   aCysGlyGlyAspCysTrpSerSerCysTrp

1573 TCTGCTCGTGC...`TGCCTCCTGCTGGAGCTGCATTATC......TGCTGG 1533
     ||||||||| ||:|||||||||||||||||||| |||||
1532 GGCTGGAGGTGTAAAACGGGGTGCTCGTCGTGCTGCTGGCCTGCTG 1483
89   ysGlyCysCysAsnCysCysTrpSerCysCysCysCysAsnCysTrp 105

1532 GGCTGGAGGTGTAAAACGGGGTGCTCGTCGTGCTGCTGGCCTGCTG 1483
     ::: |||
106 Ser.............................CysCysCysCysTrpSerCysCy 115

1482 CTGCTGCTGTAGCATGTGTGTGTCTCTGGAGTGCACTTCACTTGGCTAAACA 1433
     ||||||||| ::
115 sCysCysCysCys................................TrpLeuAsnM 122

1432 GCACTGCA 1425

         :::|||
122 etvaiAla 124

seq_documentation_block:
ID   ABB61245 standard; Protein; 517 AA.
XX
AC   ABB61245;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 10527.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US092931.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05348.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 10527; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   517 AA;

alignment_scores:
         Quality: 216.00   Length: 335
         Ratio: 1.278   Gaps: 10
Percent Similarity: 50.448   Percent Identity: 27.463

alignment_block:
US-09-664-641-10 x ABB61245

Align seq 1/1 to: ABB61245   from: 1   to: 517

768 TCTGGATTGCGTATCAGAGAAAACCAAAAAGGACGACGACCGCATTTTATCATC 817
     ||||||||| ||| ||||||| :::||| | :::||| :::
185 AsnGlyArgArgIleIleGlyIleGlnAlaThrGlyAsnSerSerValLeuA1 201

818 CTCGTCTGATTATTTATTGAAGAAGGAAGAGGAAGAGGAAGAGGAAGAGGAG 867
     :::||||||| ||| |||||||||| ||||||||||| |||
201 aAlaSerAsp.lleLysGluGluAspLleGluGluGluAspProGluGluLeu 217

```
 868 CAAGTAGAAAATGAGGAACAAATTCTGAGAATGAGGGTAGTACAGATGA 917
        :::  |||||              ::::||
 218 ...AlaGlnAlaGluAlaAspGluGlnAsp..LeuGlnAspIleGluGluGlu... 228

 918 GAAGTCAAGCCCTGCCAAGCTCTCAAGAAGGGTCCCTTCAGGTCACCAGC 967
        :::            |||||            :::||::||  :
 229 ...AlaGlnAlaGlnAlaAspGluGlnProGluThrThrAlaH  243

 968 AGTTTTCACCTAAATCCAACACTGAAAAATCTAAAGGGAAATTAAATGTTT 1017
        :::|||:::                 :::||||:::
 243 isLeuSerThrGlnValLysLeuGluLeuLysLysGluIleLysMetPro 259

1018 GAT...........       CAGTTCTTCAGATTCATCACCGGGAAAAACACAGGAGAG 1055
        |||                      :::|||:::  :::  ::|||:::
 260 AspGluGluGluGluGlnAspGluAspAspGluGluLeuGlnHisGl 276

1056 AAAATTTAAACTGGCACCCCGGCCCAAGTCCCACACAGTTACCTGCAGCAAAAC 1105
        :::|||:::            :::  :::|||:::|||:::  :::
 276 yGlnLeuGlnTyrGlnIleGluSerAlaAlaThrGlnThrAlaGlyAlaAlaT 293

1106 GCAGGCTGCCTCAGGGAAAAGGAGCCTGGGTTGATTAACTTGTGTGCCAT 1155
        :::|||:::
 293 hrSerSerAlaAlaGly........... 298

1156 CTCCCACCCGTCCCAGGTAACCATTTTGCCCCCTGAGGTTCCGGGTTAATTT 1205
        :::  |||
 299 ...............CAAAACCTCCAAACGTTCTGAAAGATCAGAAA 1249
        |||:::|||
 301 aThrAlaGlyGlyGlySerGlnGlnLeuuValGlnLeuuProAlaGlySeri 318

1250 TGATAGCTTACCTGGAGTCCAGGTCGTTACGGACACTGAGGATATTACTAAT 1299
        :::|||:::
 318 leValSerThrThrThrHis...... ArgLeuSerValProValPhrGlu 332

1300 AATGCCTGACATTCAGCACGATGAAACCGGCCATTCAAAAATGTAGCACATATCTT 1349
        :::|||        |||:::|||:::|||
 333 AlaAlaAleuValGlnLeuGlnArgQArgProValTyrIleSerAsnAlaVa 349

1350 ACAGACTCTTTCACCA........ 1365
        :::  |||:::|||
 349 lGlyGlyLeuThrAlaGlyAlaThrTyrValArgMetProAlaAlaAlaLav 366

1366 ................CCTACGAAAAATTAGAACAGCAG...... 1389
        :::
 366 alIleSerArgSerProMetThrValLeuAsnValGlnValGlnAlaLeu 382

1390 ...........GTGAATCACAGCACCAGGGCACACATTACAAAATGCCAA 1424
        :::|||:::|||
 383 ...ProAlaThrGlnLeuLeuLeuLeuHisThrGlnGlnThrProAlaAl 399

1425 TGCAGTGCTGTTTAGCCCAAGTGAAAGTGACTCCAGAGACACACATGCCTAC 1474
        :::|||:::|||
 399 aAlaAlaAlaGluGlnGlnLeuuAlaLeuuAlaLeuuGlnLnArgGlnG 416

1475 AGCAGCAGCAGCAGGCGCAGCAGCCACAGCAGCCACCGGTTTTACAC 1524
        :::|||:::|||
 416 lnGluGluGlnLnAlaGlnGlnGlnAArgGlnGlnGlnGlnGlu 432

1525 CTTCAGCCCCAGCAGGATAATGCAGCTCCAGCAGCAGCAGCAGCAGCAGAT 1574
        :::  |||:::|||
 433 GlnGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 449

1575 CTCTCAGCCAACCTTACCCCCACAGCGCCGCCGCCATCCATTTTCAGCAGC 1624
        :::|||:::|||
 449 ProLeuGlnProThrProHisProValLysHisHisSerHisSerAlaSer 462

1625 AGCAGAGCAACAGCAGCGCCACCACCAATCCGCCTCAGCAGCAATCAGCTTTTT 1674
        :::|||:::|||
 462 lnGlnGlnGlnGlnThrThrHisProValLysHisHisSerHisSerAlaSer 478

1675 GGA 1677
        |||
 479 Gly 479
```

seq_name: /SIDSS/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG27784

ID ABG27784 standard; Protein; 90 AA.
XX
AC ABG27784;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #27775.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS91971.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID No 58143; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence   90 AA;

alignment_scores:
          Quality: 215.50   Length: 107
            Ratio: 5.388   Gaps: 4
Percent Similarity: 37.383   Percent Identity: 33.645

```
alignment_block:
US-09-664-641-10/rev x ABG27784

Align seg 1/1  to: ABG27784  from: 1  to: 90

1655 GGCGATGGTGGTGGCTGCTGTTGCTGCGTGCTGCTGGCTGTGAAAAATGGATG 1606
      ||| ||||||||||||||||||||||||||||||||||||||||||||||
   7 GlyGlyGlyGlyGlyCysCysCysCysCysCys.......... 19

1605 CGGCGCGCTGCTGGGGGGCTAAGGTTGCTGAGAGACTCTGCTGCTGCTGCTT 1556
                                    |||||||||||||||||
  20 .......................................CysCysCysCysC 25

1555 GCTGGAGCTGCATTATCTGCTGGGGCTGAAGGTGTAAAACCGGGTGCTGC 1506
      ||                                        |||||||
  25 ys..........................................CysCys 27

1505 TGCTGCTGCTCCTGGGCCCTGCTGCTGCTGTAGCANGTGTGTCTGG 1456
      ||||||||||| :::||||||||||||||||||
  28 CysCysCysCysCysCysCysCysCys............... 38

1455 AGTCACTTTCACTTGGCTAAACACGCACTGCATGGGCATTTGTATGTCCCT 1406
      ||||||||||||||||||||||||||||||||||||||||||
  38 ................................................ 38

1405 GCTGGCTGTGATTCACCTGCTGT..................TCTAAATTT 1374
      ||||||                                   ||||||
  39 ............CysCysTrpLysAspLeuArgAspSerLysAla 49

1373 TTCGTAGGTGCTGAAAGAGTC 1353
      |||||:::: ::||||||
  50 PheIleSerPheSerArgVal 56
```

OM of: US-09-664-641-10 to: Pending_Patents_AA_New:* out_format : pfs

Date: Jun 4, 2002 5:45 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL -frame+_n2p.model
-Q=/cgn2_1/USPTO_spool/US09664641/runal_04062002_110933_9224/app_query.fasta_1.3697
-DB=Pending_Patents_AA_New -QPMT-lastan -SUFFIX-lastan
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH-0.100 -LOOPCL-0.000
-LOOPEXT=0.000 -QGAP0P=4.500 -QGAPEXT=0.050 -XGAP0P=10.000
-XGAPEXT=0.500 -FGAP0P=6.000 -FGAPEXT=7.000 -YGAP0P=10.000
-YGAPEXT=0.500 -DELOP-6.000 -DELEXT-7.000 -START=1
-MATRIX-blosum62 -TRANS-human40.cdi -LIST-45 -DOCALIGN-200
-THR_SCORE-pct -THR_MAX-100 -ALIGN-15 -MODE-LOCAL
-OUTFMT-pfs -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USHR=US09664641_@CCN1_1.83 -NCPU=6 -ICPU=3 -I=NOLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search Information block:
Query: US-09-664-641-10
Query length: 3580
Database: Pending_Patents_AA_New
Database sequences: 160696
Database length: 45962370
Search time (sec): 96.670000

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
/cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:US-09-300-008B-54	+	428.00	471.23	2.0e-18	84	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-381		275.00	276.58	5.3e-09	381	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-381		275.00	276.58	5.3e-09	381	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-381		275.00	276.58	5.3e-09	381	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-381		275.00	276.58	5.3e-09	381	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-381		275.00	276.58	5.3e-09	381	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-029-217A-24		272.50	268.90	8.1e-09	24	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-89		269.00	271.58	1.2e-08	89	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-149		269.00	271.58	1.2e-08	149	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-89		269.00	271.58	1.2e-08	89	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-89		269.00	271.58	1.2e-08	89	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-89		269.00	271.58	1.2e-08	89	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-89		269.00	271.58	1.2e-08	89	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-147		266.50	262.82	1.8e-08	147	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-371		261.50	259.83	3.6e-08	371	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-371		261.50	259.83	3.6e-08	371	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-371		261.50	259.83	3.6e-08	371	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-371		261.50	259.83	3.6e-08	371	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-371		261.50	259.83	3.6e-08	371	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-99		260.00	261.80	4.1e-08	99	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-99		260.00	261.80	4.1e-08	99	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-99		260.00	261.80	4.1e-08	99	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-99		260.00	261.80	4.1e-08	99	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-99		260.00	261.80	4.1e-08	99	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-15		259.00	257.38	5.0e-08	27	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-15		259.00	257.38	5.0e-08	27	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-127		257.50	257.96	6.0e-08	127	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-127		257.50	257.96	6.0e-08	127	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-127		257.50	257.96	6.0e-08	127	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-864-127		257.50	257.96	6.0e-08	127	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-127		257.50	257.96	6.0e-08	127	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-53		257.00	259.69	6.2e-08	16	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-53		257.00	259.69	6.2e-08	16	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-543		257.00	252.58	6.9e-08	3	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-543		257.00	252.58	6.9e-08	3	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-543		257.00	252.58	6.9e-08	3	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-543		257.00	252.58	6.9e-08	3	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-543		257.00	252.58	6.9e-08	3	
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-121-062-581		256.50	256.50	6.4e-08		
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-543		257.00	252.58	6.9e-08		
/cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-291		255.50	254.70	8.0e-08		

seq_name: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:US-09-300-008B-54

seq_documentation_block:
; Sequence 54, Application US/09300008B
; GENERAL INFORMATION:
; APPLICANT: Concannon et al.
; TITLE OF INVENTION: A GENE ASSOCIATED WITH NIJMEGEN BREAKAGE
; TITLE OF INVENTION: SYNDROME, ITS GENE PRODUCT AND METHODS FOR THEIR USE
; FILE REFERENCE: 9924-0003-228
; CURRENT APPLICATION NUMBER: US/09/300,008B
; CURRENT FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: US 60/083,269
; PRIOR FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 54
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-300-008B-54

alignment_scores:
Quality: 428.00 Length: 84
Ratio: 5.095 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 97.619

alignment block:
US-09-664-641-10 x US-09-300-008B-54

Align seq 1/1 to: US-09-300-008B-54 from: 1 to: 84

```
2488 GAATTGACCCCTTTTGTGCTTTTCACTGGATTCGAGCCTTGTCCAGGTTCA 2537
     ||||||||||||||||||||||||||||||||||||||||||||||||||
   1 GluLeuThrProPheValLeuPheThrGlyPheGluProCysProGlySer 17

2538 ACAGTATATTAAGAAACCTCACATTCTTGGTGGACACGTTGCGGACTCTG 2587
     |||||||||||||||||||||||||||||||||||||||||||||||||
  17 nGlnTyrIleLysLysLeuTyrIleLeuGlyGlyGluValAlaAlaGluSer 34

2588 CACAGAAGTGCACACACCTCATTGCCCAGCAAAAGTGACTCGTCGCACCGTGAAG 2637
     ||||||||||||||||||||||||||||||||||||||||:::||||
  34 laGlnLysCysThrHisLeuIleAlaSerLysValThrArgThrLeuLys 50

2638 TTCCTGGACGGCGCGCATTTCTGTCGGTGAAGGCACACTATAGTGACGGCCAGAGTGGGCT 2687
     |||||||||||||||||||||||||||||||||||||||||||||||||
  51 PheLeuLeuAspGlyAlaHisPheCysArgValLysAlaThrIleSerAspGlyGlnSer 67

2688 GGAAGAAATGCTTCAGGTGTCAGAGAGTTCATTGAGATCGAGAGCAGAGAACTACAGATTC 2737
     ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  67 uGluGluCysPheArgCysSerGlnLysPheIleAspGluGlnAsnTyrIleLeu 84

67 uGluGluCysPheArgCysSerGlnLysPheIleAspGluGlnAsnTyrIleLeu 84

2738 TC 2739
     ||
  84 eu 84
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-381

seq_documentation_block:
; Sequence 381, Application US/10123155
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang

APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
CURRENT FILING DATE: 2002-04-15
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 381
LENGTH: 2236
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-155-381

alignment_scores:
    Quality: 275.00
    Ratio: 4.231
Percent Similarity: 49.618

alignment_block:
US-09-664-641-10/rev x US-10-123-155-381

Align seq 1/1 to: US-10-123-155-381 from: 1 to: 2236

1700 GGAATCTCCACTCGTGGGATCATGTCCAAAAAGTCGATCCTTCCTGAGGCGA 1651
     ||| ||||||::::::|||||| |||||| |||
20 GlyGlyCysThrGlyAlaGlyIlyCysGlyCysGlyCysThrGlyCy 36

1650 TGGTCGGTCGGCTGCGTCGT................TGCCTGC.................. 1625
     ||||||| ||||
36 sGlyGlyCysThrGlyAlaGlyIlyCysGlyCysGlyCysThrGlyCy 53

1624 GCTGCGTCTGAAAATGGAATGCGGCGGCCTGCGTCGCTGGCGTGGGGGTAAGGTTGCTGGCGT 1575
     ||||||| ||| |||| ||| |||| |||
53 ysCysGlyCysGlyGlyCysCysGlyGlyCysThrGlyCysGlyCysGlyGly 69

1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCANTATCTGCTG 1534
     ||||| ||||||||||::::|||| ||| ::::||||
70 CysCysGlyCysAlaCysGlyCysGlySAlaAlaCysGlyCysCysGl 86

1533 GGGCTGAAGGTGTAAAACCGGGTCGTCCTGCGAGTCGCGCCTGCCT 1484
     ||||||||::::: ||| |||| |||
86 yGlyCysGlyCysGlyGlyCysGlyCysGlyCysGlyCysCysGl 103

1483 GCTGCTGCTGTAGCANTGTGTCGTCGTGGAGTCGACTTTCACTGGCTAAAC 1434
     ||||||||||: ||| :::: |||| ||||
103 ysCysGlyCysGlyyCysGlyyCysGlyCysGlyCysGlyCysC 116

1433 AGCACTGCANTGGCANTTGTATNGTCCCGCTGGCGTGGTGAGTTCACCTGCTG 1384
     ||| :::: |||
117 GlyCysGlyCysGlyCysGlyCysGlyCysCysGlyCysGlyCysCy 133

1383 TTCTAAAATTTTCCTTAGGTGCTGAAAGAGCTGTAAGATATGT 1341
     ||| :|||||| |||
133 sCysAlaGlyGlyThrGlyAlaGlyiyCysGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-381

seq_documentation_block:
Sequence 381, Application US/1037871
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.

APPLICANT: Beresini,Maureen
APPLICANT: DeForge,Laura
APPLICANT: Desnoyers,luc
APPLICANT: Filvaroff,Ellen
APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C153
CURRENT APPLICATION NUMBER: US/10/137,871
CURRENT FILING DATE: 2002-05-03
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 381
LENGTH: 2236
TYPE: DNA
ORGANISM: Homo Sapien
US-10-137-871-381

alignment_scores:
    Quality: 275.00
    Ratio: 4.231
Percent Similarity: 49.618

alignment_block:
US-09-664-641-10/re x US-10-137-871-381

Align seq 1/1 to: US-10-137-871-381 from: 1 to: 2236

1700 GGAATCTCCACTCGTCGGATCATGTCCAAAAAGTCGATCCTTCCTGAGGCGA 1651
     ||| ||||::::::|||||| |||||| |||
20 GlyGlyCysThrGlyAlaGlyIlyCysGlyCysGlyCysThrGlyCy 36

1650 TGGTCGGTCGGCTGCGTCGT................TGCCTGC.................. 1625
     ||||||| ||||
36 sGlyGlyCysThrGlyAlaGlyIlyCysGlyCysGlyCysThrGlyCy 53

1624 GCTGCTGCTGAAAATGGAATGCGGCGGCCTGCGTCGCTGGCGTGGGGGTAAGGTTGCTGGCGT 1575
     ||||||| ||| |||| ||| |||| |||
53 ysCysGlyCysGlyyGlyCysGlyGlyCysThrGlyCysGlyCysGlyGly 69

1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCANTATCTGCTG 1534
     ||||| ||||||||||::::|||| ||| ::::||||
70 CysCysGlyCysAlaCysGlyCysGlySAlaAlaCysGlyCysCysGl 86

1533 GGGCTGAAGGTGTAAAACCGGGTCGTGCCTGCGAGTCGCGCCTGCCT 1484
     ||||||||::: ||| |||| |||
86 yGlyCysGlyCysGlyGlyCysGlyCysGlyCysGlyCysCysGl 103

1483 GCTGCTGCTGTAGCANTGTGTCTCGTGGAGTCGACTTTCACTGGCTAAAC 1434
     ||||||||||: ||| :::: |||| ||||
103 ysCysGlyCysGlyyCysGlyyCysGlyCysGlyCysGlyCysC 116

1433 AGCACTGCANTGGCATTTGTATNGTCCCGCTGGCGTGGTGAGTTCACCTGCTG 1384
     ||| :::: |||
117 GlyCysGlyCysGlyCysGlyCysGlyCysCysGlyCysGlyCysCy 133

1383 TTCTAAAATTTTCGTAGGTGCTGAAAGAGCTGTAAGATATGT 1341
     ||| :|||||| |||
133 sCysAlaGlyGlyThrGlyAlaGlyiyCysGlyCysGlyCysThrCysCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-141-761-381

Left column:

```
seq_documentation_block:
; Sequence 381, Application US/10141761
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;  LENGTH: 2236
;  TYPE: DNA
;  ORGANISM: Homo Sapien
US-10-141-761-381

alignment_block:
US-09-664-641-10/rev x US-10-141-761-381

Align seq 1/1   to: US-10-141-761-381   from: 1   to: 2236

Percent Similarity:     49.618   Percent Identity: 40.458

alignment_scores:
          Quality:   275.00          Length:   131
          Ratio:       4.231         Gaps:        4
```

```
1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
       ||||                          |||||        |||
  20 GlyGlyCysThrGlyAlaGlyCysGlyCysGlyCysThrCysGlyCy 36

1650 TGGTGGTGGCTGCTGT.......TGCTGC.............TGCT 1625
       |||||||       |||||||           ||||
  36 sGlyCysGlyCysGlyCysGlyCysThrCysGlyCysCysGly 69

1624 GCTCGTCGTCAAAATGGATGCGGGGCCGGGGGTCCTGCTGCTGCTGCTG 1575
       |||||||||||||||||  ||||||||||||  ||||  |||||
  53 ysCysCysGlyCysGlyCysGlyCysCysThrGlyCysGlyCysGly 69

1574 ATCTGCTGC........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
       ||||||                            ::|||
  70 CysCysCysAlaCysGlyCysGlyCysCysAlaAlaCysCysCysGl 86

1533 GGGCTGAAGGTGTAAAACCGGGTGTGCTGCTGCTGCTGCTGGGCCTGCT 1484
       ||||                    |||||||||||||||  |||
  86 yGlyCysCysGlyCysGlyCysGlyCysCysGlyCysThrAlaGlyCysC 103

1483 GCTGCTGCTGTACCATGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
       ||||||||||||::::  |||  |||
 103 YsCysCysGlyCysGlyCysGlyCysGlyCysCysGlyCysCysCys 116

1433 AGCACTGCCATTGGCAATTGTATGTCCCTGCTGGCTGTGATTCACCTGCTG 1384
       :: :::  |||  |||
 117 GlyCysCysGlyCysGlyCysGlyCysGlyGlyCysGlyCysCysGlyCysCy 133
```

Right column:

```
1383 TTCTAAATTTTTCCTTAGGTGTGCTGCAAAGAGTCTGTAAGCAIATGT 1341
       ::::|||||||             |||  |||||
 133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysGlyCys 147

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-381

seq_documentation_block:
; Sequence 381, Application US/10140864
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C184
; CURRENT APPLICATION NUMBER: US/10/140,864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
;  LENGTH: 2236
;  TYPE: DNA
;  ORGANISM: Homo Sapien
US-10-140-864-381

alignment_block:
US-09-664-641-10/rev x US-10-140-864-381

Align seq 1/1   to: US-10-140-864-381   from: 1   to: 2236

Percent Similarity:     49.618   Percent Identity: 40.458

alignment_scores:
          Quality:   275.00          Length:   131
          Ratio:       4.231         Gaps:        4
```

```
1700 GGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGGCGA 1651
       |||                           |||||        |||
  20 GlyGlyCysThrGlyAlaGlyCysGlyCysGlyCysThrCysGlyCy 36

1650 TGGTGGTGGCTGCTGT.......TGCTGC.............TGCT 1625
       ||||||||      |||||||          ||||
  36 sGlyCysG..yGlyCysGlyCysGlyCysThrCysGlyCysGly 69

1574 ATCTGCTGC........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
       ||||||||                           ::|||
  70 CysCysCysAlaCysGlyCysGlyCysCysAlaAlaCysCysCysGl 86

1533 GGGCTGAAGGTGTAAAACCGGGTGTGCTGCTGCTGCTGCTGGGCCTGCT 1484
       ||||                    |||||||||||||  |||
  86 yGlyCysCysGlyCysGlyCysGlyCysCysGlyCysThrAlaGlyCysC 103

1483 GCTGCTGCTGTGTAGCACATGTGTGTCTCTGGAGTCACTTTCACTTGGCTAAAC 1434
       ||||||||||::: |||  ::||||
 103 YsCysCysGlyCysGlyCysGlyCysGlyCysCysGlyCysCys 116
```

```
1433 AGCACTGCATTGGCANTTGTATGTCCCTGCTGCTGTGATTCACCTGCTG 1384
     ::: :::   ||| |||
117 GlyCysGlyCysCysCysGlyCysCysGlyCysGlyCysGlyCysCy 133

1383 TTCTAAATTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
     ::|||||||||       |||       |||
133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-381

seq_documentation_block:
; Sequence 381, Application US/10140923
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 381
; LENGTH: 2236
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-381

alignment_scores:
            Quality: 275.00   length: 131
            Ratio: 4.231   Gaps: 4
Percent Similarity: 49.618   Percent Identity: 40.458

alignment_block:
US-09-664-641-10/rev x US-10-140-923-381

Align seq 1/1   to: US-10-140-923-381   from: 1 to: 2236 .

```
1700 GGAACTCCACTGCTGGATCATCTCCAAAAAGCTGATGCTGTGAGGCGA 1651
     ||:|||| |  || |||   |||       |||    |||
20 GlyGlyCysThrGlyAlaGlyCysGlyCysGlyCysThrCysGlyCy 36

1650 TGGTGGTGGCGTGCTGT................TGCTGC 1625
     |||||||| ||||                   |||
36 sGlyGlyCysGlyCysGlyCysGlyCysGlyCys.....TGCTGC 53

1624 GCTCGCTGCGAAAATGGATCGCGCGCGCCTGCTGCGGGGGTAAGGTTGCTCAGAG 1575
     ||||||| |||  |||  |||  |||   |||       ||||||
53 ysCysGlyCysGlyCysThrCysGlyCysGlyCysGlyCysGlyCysGly 69

1574 ATCTGCTGC.........TGCTGCTGCTGCTGGAGCTGCATTATCTGCTG 1534
     ||||||||||         ::|||
70 CysCysCysAlaCysGlyCysCysGlyCysAlaCysGlyCysGlyCysGl 86

1533 GGGTGAAGGTGTAAAAACCGGGTTGCTGCTCCTGCTGCTGCGTGGGCCTGCT 1484
     |||      |||                 ::|||
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-029-217A-24

```
86 yGlyGlyCysCysGlyCysGlyCysGlyCysCysGlyCysCysGlyCysThrAlaGlyCysC 103
   ||| |||       |||:::       |||          |||
1483 GCTCGCTGCTGCTGTACCATCGTGTCTCTCGGAGTCACTTTCACTTGGCTAAAC 1434
     |||||||:::     |||    :::|||       |||||
103 ysCysCysCysGlyCysGlyCysGlyCysGlyCysGlyCysCys.......... 116

1433 AGCACTGCATTGGCANTTGTATGTCCCTGCTGCTGTGATTCACCTGCTG 1384
     ::: :::   ||| |||
117 GlyCysGlyCysCysCysGlyCysCysGlyCysGlyCysGlyCysCy 133

1383 TTCTAAATTTTCGTAGGTGCTGAAAGAGTCTGTAAGATATGT 1341
     ::|||||||||       |||       |||
133 sCysAlaGlyGlyThrGlyAlaGlyCysGlyCysThrCysCys 147
```

seq_documentation_block:
; Sequence 24, Application US/10029217A
; GENERAL INFORMATION:
; APPLICANT: OLSON, ERIC N.
; APPLICANT: WANG, DA-ZHI
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO A CARDIAC-SPECIFIC
; TITLE OF INVENTION: NUCLEAR REGULATORY FACTOR
; FILE REFERENCE: UTSD:6955US
; CURRENT APPLICATION NUMBER: US/10/029,217A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/257,761
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 3907
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-029-217A-24

alignment_scores:
            Quality: 272.50   Length: 169
            Ratio: 3.406   Gaps: 7
Percent Similarity: 47.337   Percent Identity: 34.320

alignment_block:
US-09-664-641-10/rev x US-10-029-217A-24

Align seq 1/1   to: US-10-029-217A-24   from: 1 to: 3907 .

```
1778 GCCAGCAGTTGCTTATCAGACATCTGCTCTGGATTAATCCGCAATTGCAAA 1729
     |||:|||  |||   :::      |||          :::|||
2317 AlaThrCysCysCysThrGlyCysCysAlaGlyGlyValGlyVala...... 2330

1728 CACACACATCCCAATAAGAGAGCCTTCTCTCGGAATCCCACTGCTGCGATCGT 1679
     |||||||          |||                   :::|||||
2331 ................AlaGlyValAlaAlaGlyCysC 2340

1678 GTCCAAAAAGCTGATGCTGTGAGGCGATGGT... 1647
     ||          |||  |||
2340 ysAlaThrCysCysCysGlyCysAlaAlaGlyAlaCysAlaGlyThrCys 2356

1646 ..........GGTGGTGGCTGCTGTTGCTGCTTGCTGCTGGTGGTGAAAATGGATG 1606
     |||||||   ||||||||||||||
2357 ThrGlyThrGlyAlaGlyThrCysCysCysCysGlyCysThrGlyGlyCy 2373

1605 CGGCGCGGCGTGCTGGGGTAAGGTTGCTGCTGAGAGATCGTCGTCCTGCTGCGCT 1556
     |::|||||||   |||||||                ::||||||
2373 sAlaGlyCysAlaCysAlaGlyCys.......CysAlaThrCysAlaCysC 2388

1555 GCTGGAGCTGCATTATCTGCTGGGGGCGTGAAGGTTAAAACCGGGTGCTGC 1506
     ||  ||||| :: |||       |||
2388 ysThrThrCysThrGlyAlaGlyCysThrCysCysGlyCysCysCys 2404

1505 TGC................TGCTGCTGCTGGGGCCTGCTGCTGCTGCTG 1474
     |||                ||||||||    |||||
```

Left column

```
2405 CysAlaGlyGlyCysThrGlyCysCysThrCysCysAlaCysCysThrCysCy 2421
     |||:|||||  |:  :||
1473 TACCATGTGTCTCTCGACHCACHTCACHTGGCTAAACAGCACHTGCAT 1424
2421 sThrCysAlaGlyCysThr........   |||   ........CysAlaC 2432
     :|:  |||
1423 TGGCATTTGTTATGTCCCTGCCTGGGTGTGTGAATTCACCTGCTGTTCTAAATTT 1374
2432 ysCysCysThrCysCysCys........   |||   ThrCysCysThrGly 2444
     :||  ||||||||
1373 TTCCGTA............   ||   GGTCCTGAAAGAGTCTGTAAGATATG 1342
2445 GlyAlaCysGlyCysCysThrGlyGlyAlaGlyGlyAlaCysThrThrCy 2461
     :||:  :|||  |||
1341 TGGTACA 1335
2461 sCysThr 2463
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-89

seq_documentation_block:
; Sequence 89, Application US/10123155
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-89

alignment_block:
US-09-664-641-10/rev x US-10-123-155-89

alignment_scores:
      Quality: 269.00   Length: 168
      Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

Align seg 1/1   to: US-10-123-155-89   from: 1   to: 1893

```
1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATCGTGCTGAGG 1654
     :|||:  ||||||||  ::||  |||:
60   AlaGlyAlaCysThrThrGlyAlaCysThr........CysCysCysGl 73
1653 CGATGGTGGTGCTGCTGCTGTTGTCGTGCGC........... 1629
     ||||||||||||||||  ::||
73   yCysGlyCysGlyCysCysAlaAlaCysCysCysThrGlyCysT 90
1628 ............TGCTGCTGC 1620
```

Right column

```
90   hrThrAlaThrCysCysCysThrGlyAlaCysCysGlyThrCysGly 106
     |||||||||
1619 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaAlaThrCysThrGlyCy 1606
     :|:||||  ......TGTGAAAATGGATG
107  AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaAlaThrCysThrGlyCy 123
     ...
1605 CGGCGGCTGCTGGGGGGTTAAGGTTGCTGACAGANCTGCTGTGCTGC.........T 1562
123  sAlaGlyCysGlyCysCysGlyAlaGlyThrCysGlyCysGlyCysGlyC 140
1561 GCTGCTGCTGGAGCTGCAATTATCTGCTGGGGGCTGAAGGTGTAAAACCGGG 1512
140  ysCysGlyCysGlyCysCysThrCysGlyCysCys............CysGly 150
1511 TGCTGCTGCTGCTGCTGCTGGGCCTGTCCTGC...TGCTGCTGTAGCCATGTG 1465
151  CysCysGlyCysCysCysCysAlaCysGlyCysCysCysAlaCysGlyCysCysThrCysCy 167
1464 TGTCTCTGGAGTCACT...TTCACTTGGGTAAACAGCACTGCATTGGGCAT 1418
167  sThrGlyCysGlyCysThrCysThrThrCysCysThrThrGlyThrThrThrThrT 184
1417 TTGTATGTCCCTGCTGGCTGTGTGAATTCACCTGCTGTTCTAAATTTTTCGTA 1368
184  hrAlaCysThrCys............CysThrCysCysThrThrThrThrCys 197
1367 GGTGCTGCAAAGAGTCTGTAAGATATGTGTCCTACATTTGATGGCCGGTCAT 1318
198  AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214
1317 CTGC 1314
214  aCys 215
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-545-149

seq_documentation_block:
; Sequence 149, Application US/10063545
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,545
; CURRENT FILING DATE: 2002-05-02
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 149
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-545-149

alignment_block:
US-09-664-641-10/rev x US-10-063-545-149

alignment_scores:
      Quality: 269.00   Length: 168
      Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

Align seg 1/1   to: US-10-063-545-149   from: 1   to: 1893

Left column

alignment_scores:

```
1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
        ::|||  |||::|||:::|||::|||
  60 AlaGlyAlaCysThrThrGlyAlaCysThr.......CysCysCysGl 73

1653 CGATGGTGGTGGCTGCTGCTGTTGCTCC.................... 1629
        |  |||||||||||||||
  73 yCysGlyCysGlyCysGlyCysCysGlyCysCysAlaaCysCysCysThrGlyCysT 90

1628 ...........TGCTGCTGC............ 1620
              |||||||||
  90 hrThrAlaThrCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

1619 ...........................TGTGAAAATGGATG 1606
                                  ||  |||||||
 107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysCysThrGlyCy 123

1605 CGGCGGCCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC.....T 1562
        |:::||||||     |||      |||           ||||||||    |
 123 sAlaGlyCysGlyCysGlyCysAlaGlyThrCysCysCysGlyGlyC 140

1561 GCTGCTGCTGGAGCTGCATTATCTGCtGGG. . :TGAAGGTGTAAAACCGGG 1512
        |||||||||  ||||||||          |||           |
 140 ysCysCysCysThrCysThrGlyGGGC::TGAAGGTGTAAAACCGGC 150

1511 TGCTGCTGCTGCTCGGGGCCTGCTGC... TGCTGCTGTAGCATGTG 1465
        ||||||||||::|||    ||||||||:::           ||
 151 CysCysCysCysAlaCysAlaCysCysCysThrGlyCy 167

1464 TGTCTCTGGAGTCACT... TTCACTTGGCTAAAACAGCACTGCATTGGCAT 1418
        ::::|||:|||::||           |||:||||||:::           ::
 167 sThrGlyGlyCysThrThrCysThrThrCysGlyThrGlyGlyGlyC 184

1417 TTGTATGTCCCTGCCTGGCTGTGATTCACCTGCGTTCTAAATTTTTCGTA 1368
        ::::|||:|||              |||||               ::
 184 hrAlaCysThrCys.......... ...CysThrCysCysThrThrThrThrCys 197

1367 GGTGCTGAAAGAGTCTGTTGAAGATATGTCCTACATTTGATGGCCGGTTCAT 1318
        ::::|||||                |||:||||||:::           |||
 198 AlaThrThrCysAlaThrAlaAlaLaAlaAlaLaAlaGlyCysThrAl 214

1317 CTGC 1314
        |||
 214 aCys 215
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-063-685-149

seq_documentation_block:
; Sequence 149, Application US/10063685
; GENERAL INFORMATION:
; APPLICANT: Eaton,Dan L.
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gerrilsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230R1C1
; CURRENT APPLICATION NUMBER: US/10/063,685
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 149
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-063-685-149

Right column

Quality: 269.00   Length: 168
Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

alignment_block:
US-09-664-641-10/rev x US-10-063-685-149

Align seg 1/1 to: US-10-063-685-149 from: 1 to: 1893

```
1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
        ::|||  |||::|||:::|||::|||
  60 AlaGlyAlaCysThrThrGlyAlaCysThr.......CysCysCysGl 73

1653 CGATGGTGGTGGCTGCTGCTGTTGCTCC.................... 1629
        |  |||||||||||||||
  73 yCysGlyCysGlyCysGlyCysCysGlyCysCysAlaaCysCysCysThrGlyCysT 90

1628 ...........TGCTGCTGC............ 1620
              |||||||||
  90 hrThrAlaThrCysCysThrThrGlyAlaCysCysGlyThrCysGly 106

1619 ...........................TGTGAAAATGGATG 1606
                                  ||  |||||||
 107 AlaGlyThrGlyThrCysAlaGlyAlaGlyAlaThrCysCysCysThrGlyCy 123

1605 CGGCGGCCTGCTGGGGGTAAGGTTGCTGAGAGATCTGCTGCTGC.....T 1562
        |:::||||||     |||      |||           ||||||||    |
 123 sAlaGlyCysGlyCysGlyCysAlaGlyThrCysCysCysGlyGlyC 140

1561 GCTGCTGCTGGAGCTGCATTATCTGCtGGG. . :TGAAGGTGTAAAACCGGC 1512
        |||||||||  ||||||||          |||           |
 140 ysCysCysCysThrCysThrGlyGGGC::TGAAGGTGTAAAACCGGC 150

1511 TGCTGCTGCTGCTCGGGGCCTGCTGC... TGCTGCTGTAGCATGTG 1465
        ||||||||||::|||    ||||||||:::           ||
 151 CysCysCysCysAlaCysAlaCysCysCysThrGlyCy 167

1464 TGTCTCTGGAGTCACT... TTCACTTGGCTAAAACAGCACTGCATTGGCAT 1418
        ::::|||:|||::||           |||:||||||:::           ::
 167 sThrGlyGlyCysThrThrCysThrThrCysGlyThrGlyGlyGlyC 184

1417 TTGTATGTCCCTGCCTGGCTGTGATTCACCTGCGTTCTAAATTTTTCGTA 1368
        ::::|||:|||              |||||               ::
 184 hrAlaCysThrCys.......... ...CysThrCysCysThrThrThrThrCys 197

1367 GGTGCTGAAAGAGTCTGTTGAAGATATGTCCTACATTTGATGGCCGGTTCAT 1318
        ::::|||||                |||:||||||:::           |||
 198 AlaThrThrCysAlaThrAlaAlaLaAlaAlaLaAlaGlyCysThrAl 214

1317 CTGC 1314
        |||
 214 aCys 215
```

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-137-871-89

seq_documentation_block:
; Sequence 89, Application US/10137871
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerrilsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K

; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C153
; CURRENT APPLICATION NUMBER: US/10/137,871
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-89

alignment_scores:
        Quality: 269.00   Length: 168
        Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

alignment_block:
US-09-664-641-10/rev x US-10-137-871-89

Align seg 1/1   to: US-10-137-871-89   from: 1 to: 1893 ..

1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
     ::|||| |||::||:||:|||:||
  60 AlaGlyAlaCysThrThrGlyAlaCysThr........CysCysCysGl 73

1653 CGATGGTGGTGGCGTGCCTGTTGCTGC......................... 1629
     ||| |||||||||||||||
  73 yCysGlyCysGlyCysCysThrCysAlaAlaCysCysCysThrGlyCysT 90

1628 ....TGCTGCTGC................... 1620
                     |||||||||
  90 hrThrAlaThrCysCysThrThrGlyAlaThrCysCysGlyThrCysGly 106

1619 ....TGTGAAAATGGATG 1606
                            ||| ::||:||||
 107 AlaGlyThrGlyThrCysAlaLaGlyAlaLaGlyAlaLaThrCysCysThrGlyCy 123

1605 CGGCGGCTGCTGGCGTAAGGTTGCTGAGAGATCTGCTCGC......T 1562
     |::|||||| ||||||||||
 123 sAlaGlyCysCysGlyCysCysGlyAlaGlyThrCysCysGlyGlyGlyc 140

1561 GCTGCTGCTGGAGCTGCATTATCTGCTGCGGGGCTGGAAGGTGTAAAACGGG 1512
     |||||||| |||||||||
 140 ysCysGlyCysGlyCysCysThrCysCys..................... 150

1511 TGCTGCTGCTGGTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
     |::||||||||||:::|| :||||||||:::
 151 CysCysCysAlaCysCysAlaCysCysCysCysThrCysCy 167

1464 TGTGTCTGGAGTGCACT.....TTCACTTGGCTAAACAGCACTGCAATTGGCAT 1418
     |:::::::|||| :::||||:::||||
 167 sThrGlyGlyCysThrCysThrThrCysThrCysGlyThrThrThrThrT 184

1417 TTGTATGTCCCTGCCTGGCTGTGGATTCACCTCGTTCTAAAATTTTCGTA 1368
     :|||| |||||||||:::
 184 hrAlaCysThrCys........CysThrCysCysThrThrThrThrThrCys 197

1367 GGTGCTGAAAGAGTCTGTAAGAATATATGTCTACACATTTGATGGCCGGTTCAT 1318
     :||:|||| |||||||||:::
 198 AlaThrThrCysAlaThrAlaLaCysAlaLaAlaAlaAlaGlyCysThrAl 214

1317 CTGC 1314
     |||
 214 aCys 215

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_CCmB.pep:US-10-141-761-89

seq_documentation_block:

; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C198
; CURRENT APPLICATION NUMBER: US/10/141,761
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-761-89

alignment_scores:
        Quality: 269.00   Length: 168
        Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

alignment_block:
US-09-664-641-10/rev x US-10-141-761-89

Align seg 1/1   to: US-10-141-761-89   from: 1 to: 1893 ..

1703 TCTGGAATCTCCACTGCTGGATCATGTCCAAAAAGCTGATGCTGCTGAGG 1654
     ::|||| |||::||:||:|||:||
  60 AlaGlyAlaCysThrThrGlyAlaCysThr........CysCysCysGl 73

1653 CGATGGTGGTGGCGTGCCTGTTGCTGC......................... 1629
     ||| |||||||||||||||
  73 yCysGlyCysGlyCysCysThrCysAlaAlaCysCysCysThrGlyCysT 90

1628 ....TGCTGCTGC................... 1620
                     |||||||||
  90 hrThrAlaThrCysCysThrThrGlyAlaThrCysCysGlyThrCysGly 106

1619 ....TGTGAAAATGGATG 1606
                            ||| ::||:||||
 107 AlaGlyThrGlyThrCysAlaLaGlyAlaLaGlyAlaLaThrCysCysThrGlyCy 123

1605 CGGCGGCTGCTGGCGTAAGGTTGCTGAGAGATCTGCTCGC......T 1562
     |::|||||| ||||||||||
 123 sAlaGlyCysCysGlyCysCysGlyAlaGlyThrCysCysGlyGlyGlyc 140

1561 GCTGCTGCTGGAGCTGCATTATCTGCTGCGGGGCTGGAAGGTGTAAAACGGG 1512
     |||||||| |||||||||
 140 ysCysGlyCysGlyCysCysThrCysCys..................... 150

1511 TGCTGCTGCTGGTGCTGCTGGGCCTGCTGC...TGCTGCTGTAGCATGTG 1465
     |::||||||||||:::|| :||||||||:::
 151 CysCysCysAlaCysCysAlaCysCysCysCysThrCysCy 167

1464 TGTGTCTGGAGTGCACT.....TTCACTTGGCTAAACAGCACTGCAATTGGCAT 1418
     |:::::::|||| :::||||:::
 151 CysCysCysAlaCysCysAlaCysCysCysCysThrCysCy 167

167 sThrGlyGlyCysThrCysThrCysCysThrGlyThrThrThrThrT
1417 TTGTATGTCCCTGCCTGCCTGCCTGATTCACCTGCGTTCTAAATTTTCGTA 1368
     ||| |||
184 hrAlaCysThrCys........................CysThrCysCysThrhrThrThrCys 197
     |||||||||||::|
1367 GGTGCTGAAAAGAGTCTGTAAGATATGTGCTACATTGATGGGCCGGTCAT 1318
     :::::
198 AlaThrThrCysAlaThrAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214
     |||
1317 CTGCC 1314
     |||
214 aCys 215

seq_documentation_block:
; Sequence 89, Application US/10140864
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P333ORlC184
; CURRENT APPLICATION NUMBER: US/10/140, 864
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-864-89

alignment_scores:
     Quality: 269.00   Length: 168
     Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

Align seq 1/1 to: US-10-140-864-89 from: 1 to: 1893

alignment_block:
US-09-664-641-10/rev x US-10-140-864-89 ..

1703 TCTGGAATCTCCACTGCTGGAATCANTCTCAAAAAGCTGATGCTGCTGAGG 1654
     |||:||| |||::||:::||| |||
60 AlaGlyAlaCysThrThrGlyAlaCysThr........................CysCysCysGl 73
     |||||
1653 CGAtGGTGGTGCGTGGCGTGCGTGTTCCGTGCC........................ 1629
     ||| |||||||||||||||||||:
73 yCysGlyCysGlyCysGlyCysSCysCysGlyCysThrGlyCysT 90
     |||:||:||||||
1628 ....TCCTGCTGCC........................ 1620
     |||||||||
90 hrThrAlathrCysCysGlyThrThrGlyAlaCysGlyThrCysGly 106
     ..............................TGTGAAAATGGATG 1606
1619

---

107 AlaGlyThrGlyThrCysAlaGlyAlaGlylyAlaThrCysCysThrGlyCy 123
     ||| :::|||||
1605 CGGCGGCGCTGCTGCGGGGGTAAGGTTTCTGACAGAGATCTCCTCTGCC........................T 1562
     |:::|||||| |||||
123 sAlaGlyCysGlyCysGlyCysGlyAlaGlyThrCysGlyCysGlyGlyC 140
     |||||||||:
1561 GCTGCGTGCTGGAGCTGCATTATCTGCCTGGGGGCTGAAGGTGTAAAACCGGG 1512
     :::|||| |||||||||| |||
140 yaCysGlyCysGlyCysThrCysThrCysCys........................CysGly 150
     |||
1511 TGCTGCGTGCTGCTGCTGCTGCGGCCGCCTGC........................TGCTGCTGTAGCATGTG 1465
     |||||||||||:||| ||||||||||::: ||
151 CysCysGlyCysCysAlaCysGlyCysGlyCysAlaCysGlyCysThrCysGy 167
     |:::|||||| |||:: ::
1464 TGTCTCTGGAGTCACT...TTCACTTGGCTAAACACCACTGCATTGGCAT 1418
     |:::|||||||||| ||| :::
167 sThrGlyGlyCysGlyThrThrThrThrCysCysThrGlyGlyThrThrThrT 184
     |:::||||||||
1417 TTGTATGTCCCTGCCTGCCTGCCTGATTCACCTGCGTTCTAAATTTTCGTA 1368
     |||||||| |||:::
184 hrAlaCysThrCys........................CysThrCysCysThrhrThrThrCys 197
     |||||||||||::|
1367 GGTGCTGAAAAGAGTCTGTAAGATATGTGCTACATTGATGGGCCGGTCAT 1318
     ::::
198 AlaThrThrCysAlaThrAlaAlaAlaCysAlaAlaAlaAlaGlyCysThrAl 214
     |||
1317 CTGCC 1314
     |||
214 aCys 215

seq_documentation_block:
; Sequence 89, Application US/10140923
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P333ORlC188
; CURRENT APPLICATION NUMBER: US/10/140,923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 89
; LENGTH: 1893
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-89

alignment_scores:
     Quality: 269.00   Length: 168
     Ratio: 3.494   Gaps: 8
Percent Similarity: 45.833   Percent Identity: 33.333

alignment_block:

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-923-89

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-140-864-89

seq_name: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:US-10-123-155-147

seq_documentation_block:
Sequence 147, Application US/10123155
GENERAL INFORMATION:

; APPLICANT: Baker,Kevin P.
; APPLICANT: Beresini,Maureen
; APPLICANT: Deforge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30

; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 147
; LENGTH: 3734
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-123-155-147

alignment_block:
US-09-664-641-10/rev x US-10-123-155-147

alignment_scores:
        Quality: 266.50   Length: 251
        Ratio: 2.562   Gaps: 14
Percent Similarity: 41.434   Percent Identity: 28.287

Align seq 1/1 to: US-10-123-155-147 from: 1 to: 3734 ..

1920 TGCAGTAACACAICTCTTTCTTCTTTAGTTGCCTGTGCATACGCGCTGC 1871
     |||:|::|||:|||||::
2851 CysGlyThrThrAla.............CysCysCysCysAlaTh 2862

1870 TGACTTGCACTCTCACAGAGAGAGGTGCGTGCCATCGAACTCGTGAAGGTGGGG 1821
     : |||||||:
2862 rGlyCysThr.............GlyCysThr 2868

1820 TCAACAGTGCCCCGCCATGTGCCCTGGAATTATCCTTTCCAGGTGGCCAGCAG 1771
     |||||||||| |||| |||:::||
2868 hrGlyGlyCysGlyCysAlaThrCysCysThrCysCysAlaCysCysAlaGlyGly.AlaAl 2884

1770 T..........TGCTTATCAGACATCGTCCTCTGGATAATCCGCAAITTG 1733
     |::: | |||:::||
2884 aGlyAlaGlyAlaGlyGlyCysGlyCysThrGlyThrThrCysThrCysGlyGlyCysGlyGl 2901

1732 CAAACACACATCCCAAATAAAGAAGCCTTCTTCTGGAATCTCCACTGCCTGGA 1683
     ||||:|| ||||:|||
2901 CysGlyCysThrGlyGlyGlyGlyThrGlyGl 2914

1682 TCAGTGTCCAAAAAGGCTGATGCTGCTGCTGAGGCGGATGGT.......... 1647
     |||:|| ||||
2914 AlaCys...........CysCysThrGlyGlyGlyGlyThrGlyGl 2926

1646 ....GGTGGCTGC................ 1637
     |||||||||
2926 yAlaCysAlaGlyAlaGlyAlaGlyGlyCysGlyThrCysGlyThrCysGlyCysAlaAlAa 2943

1637 ..........TGTTGCTGCTGC..........TGCTGCTGT 1617
     ||||||||| |||||||||
2943 laCysGlyCysGlyCysAlaCysThrCysCysGlyThrThrCysGlyCysCys 2959

1616 GAAAATGA............ 1597
     ::|||| |||
2960 GlyThrGlyThrGlyThrCysGlyThrThrCysThrThrCysGlyCysCysThrGlyCyCy 2976

1596 CTGGGGGTGAAGGTTGGTCAGAGAGATCTGCTGCTCGCTGC...... 1560
     ||| ::|| ||||||
2976 scysAlaAlaGlyCysCysThrCysCysCysThrGlyCysCysA 2993

1559 ..........TGCTGCTGGGAGCTGCATTATCTGCTGGGGCGTGAAGGTGTGAAA 1518
     ||||||||| ||||||||||
2993 laAlaGlyCysCysThrCysGlyThrGlyThrCysCysGlyCysCysCys 3009

1517 ACC.......... 1486
     ||| |||
2993 ...GGGTGCTGCTGCTGCTGCTGCGCTCGGGCCTG 3009

1485 CTGCTCCTGCTGCTTGTGCATGTGTGTCTCGCGGAGTCACTTTCACTTGGCTAA 1436
     ||||||||||| |:|||| :
3026 ScysCysCysCysAlaCysAlaLaCysGlyThrCysGlyThrCysGlyCysAlaTaT 3043

```
1435 ACAGCACTGCATTGGCATTTGTTATGTCCCTGCCTGGCTGTGATTCACCTGC 1386
     :: |||::: ::::||| ||| |||
3043 hrCysThrGlyCysCysThrCysCys............CysCys 3055

1385 TGTTCTAAAATTTTTTCGGTACGTGCTGAAAGACTCTCTAAGATATGTCCTAC 1336
     ||||:::  |||:::||  |||:::||
3056 CysThr...........GlyGlyCysCysCysCysAlaCysThrTh 3068

1335 A 1335

3068 т 3068
```